20

PK
505
L84E
V.20

ENDEAVOUR

VOL. XX

1961

R. BOSE

077826

गुरुकुल कांगड़ी विश्वविद्यालय, हरिद्वार
पुस्तकालय

R
विषय संख्या 505
पुस्तक संख्या L84E
V. 20
आगत पंजिका संख्या 63440

पुस्तक पर किसी प्रकार का निशान लगाना वर्जित है। कृपया १५ दिन से अधिक समय तक पुस्तक अपने पास न रखें।

63440

505
L84E
V. 20

पुस्तकालय
गुरुकुल कांगड़ी विश्वविद्यालय, हरिद्वार

COMPILED

पुस्तक-वितरण की तिथि नीचे अंकित है। इस तिथि सहित १५वें दिन तक यह पुस्तक पुस्तकालय में वापिस आ जानी चाहिए। अन्यथा ५ पैसे प्रतिदिन के हिसाब से विलम्ब-दण्ड लगेगा।

STOCK VERIFICATION
Central
Library
2024-25 G.K.D.U. HW.

५११·
१३

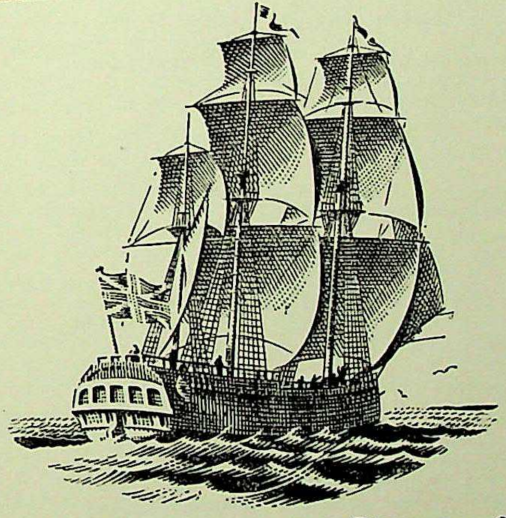
CHECKED 1973
1-iii 1 81
COMPILED

~~63440~~

ENDEAVOUR

A quarterly review, published in five languages, designed to record
the progress of the sciences in the service of mankind

077826
077826



[illegible]क प्रमाणीकरण १२८४-१२८५

VOLUME XX · 1961

AUTHOR INDEX

Abraham, E. P. and Newton, G. G. F. New penicillins, cephalosporin C, and penicillinase. 92.
Barker, G. C. and Milner, G. W. C. Polarography. 26.
Bondi, H. Gravitational waves. 121.
Bullard, Sir Edward. The Mohole. 188.
Butler, C. G. The efficiency of a honeybee community. 5.
Djerassi, Carl. Optical rotatory dispersion. 138.

Editorial
    An international council of sciences. 175.
    Investing in research. 119.
    Jubilee of the *Max-Planck-Gesellschaft*. 59.
    The ocean depths. 3.
Eyles, V. A. Sir James Hall, Bt. (1761–1832). 210.
Farthing, T. W. *See* McQuillan, M. K. and Farthing, T. W.
Folkers, K. *See* Wagner, A. F. and Folkers, K.
Glueckauf, E. Isotope separation by reversible chemical processes. 42.
Goreau, T. Problems of growth and calcium deposition in reef corals. 32.
Heymann, F. F. Positronium. 225.
Kursanov, A. L. The transport of organic substances in plants. 19.
Lewin, G. Ultra-high vacuum. 85.
McQuillan, M. K. and Farthing, T. W. Beryllium. 11.
Massey, Sir Harrie. Space research. 68.
Milner, G. W. C. *See* Barker, G. C. and Milner, G. W. C.
Nairn, R. C. Fluorescent protein tracing and the fluorescent antibody method. 78.
Newton, G. G. F. *See* Abraham, E. P. and Newton, G. G. F.
O'Connell, D. J. K. The green flash and kindred phenomena. 131.
Partington, J. R. Greek fire. 162.
Pirenne, M. H. Light quanta and vision. 197.
Pye, J. D. Echolocation by bats. 101.
Rattee, I. D. Reactive dyestuffs for cellulose. 154.
Sirlin, J. L. The nucleolus of the cell nucleus. 146.
Sjörs, H. Surface patterns in Boreal peatland. 217.

Special Article
    The beginnings of phosphorus manufacture. 40.
Stone, F. G. A. Electron-deficient compounds. 61.
Wagner, A. F. and Folkers, K. The organic and biological chemistry of mevalonic acid. 177.

BOOK REVIEWS

Acheson, R. M. An Introduction to the Chemistry of Heterocyclic Compounds. 113.
Andrade, E. N. da C. A Brief History of the Royal Society. 54.
Ashby, W. Ross. Design for a Brain (second edition, revised). 52.
Austin, C. K. (Editor). Journal of Reproduction and Fertility, Vol. I, No. I, February 1960. 52.

Baker, R. M. L. Jnr. and Makemson, M. W. An Introduction to Astrodynamics. 167.
Barker, T. C. Pilkington Brothers and the Glass Industry. 170.
Bass, A. M. and Broida, H. P. (Editors). Stabilization of Free Radicals at Low Temperatures. 112.
Berry, A. J. Henry Cavendish: His Life and Scientific Work. 54.
Bird, R. B., Stewart, W. E., and Lightfoot, E. N. Transport Phenomena. 232.

Chapman, V. J. Salt Marshes and Salt Deserts of the World. 51.
Clarke, N. (Editor). A Physics Anthology. 50.
Coates, G. E. Organo-Metallic Compounds (second edition). 168.

Dean, J. A. Flame Photometry. 232.
de Beer, Sir Gavin. The Sciences were never at War. 53.
Derry, T. K. and Williams, Trevor I. A Short History of Technology, from Earliest Times to A.D. 1900. 54.
Djerassi, C. Optical Rotatory Dispersion. 169.
Driver, J. E. (revised by). Bentley and Driver's Textbook of Pharmaceutical Chemistry (seventh edition). 53.

Emeléus, H. J. and Anderson, J. S. Modern Aspects of Inorganic Chemistry (third edition). 51.

Falconer, D. S. Introduction to Quantitative Genetics. 113.
Fierz, M. and Weisskopf, V. F. (Editors). Theoretical Physics in the Twentieth Century. 232.
Finkbeiner, Daniel T. II. Introduction to Matrices and Linear Transformations. 167.

Gilmour, D. The Biochemistry of Insects. 233.
Grassé, P.-P. (Editor). Traité de zoologie, Vol. V, Part 2, Bryozaires, Brachiopodes, Chétognathes, Pogonophores, Mollusques. 234.
Greenberg, D. M. (Editor). Metabolic Pathways, Vol. I. (Second edition of 'Chemical Pathways of Metabolism'.) 233.

Hagerup, O. and Peterson, V. Botanisk Atlas, Vol. II (translated by Carter, H. Gilbert). 234.
Harper, J. L. (Editor). The Biology of Weeds. 52.
Harris, H. Human Biochemical Genetics. 52.
Hartley, Sir Harold (Editor). The Royal Society: its Origins and Founders. 54.
Horsfall, J. G. and Dimond, A. E. (Editors). Plant Pathology, Vol. II, The Pathogen. 52.

Jones, H. The Theory of Brillouin Zones and Electronic Structure in Crystals. 112.

Kamlet, M. J. (Editor). Organic Electronic Spectral Data, Vol. I (1946–52). 233.
Klingsberg, E. (Editor). Pyridine and its Derivatives, Part I, by Barnes, R. A., Brody, F., and Ruby, P. R. 113.

Landau, L. D. and Lifshitz, E. M. Electrodynamics of Continuous Media (translated from the Russian by Sykes, J. B. and Bell, J. S.). 167.
Landsberg, P. T. Monographs in Statistical Physics and Thermodynamics, Vol. II, Thermodynamics. 232.
Lewis, J. J. Introduction to Pharmacology. 53.

McGraw-Hill Encyclopaedia of Science and Technology, Vols. I–XIV, Aba–Zyt. 166.
Manske, R. H. F. The Alkaloids: Chemistry and Physiology, Vols. VI and VII. 113.

Nord, F. F. (Editor). Advances in Enzymology, Vol. XXII. 169.

Osman Hill, W. C. Primates, Comparative Anatomy and Taxonomy, Vol. IV, Cebidae, Part A. 170.

Partington, J. R. A History of Greek Fire and Gunpowder. 53.
Pascal, P. (Editor). Nouveau traité de chimie minérale, Vol. XV, Uranium et transuraniens, Part I, Uranium. 113.
Perucca, E. Fisica Generale e Sperimentale (seventh edition, revised) Vol. I, and Vol. II Parts I and II. 167.
Plešek, J. and Zobáčová, A. Preparativní reakce v organické chemii. V. Aldolisace a příbuzné reakce. (Reactions in Preparative Organic Chemistry. Vol. V, Aldol and Similar Reactions.) 168.
Porter, C. L. Taxonomy of Flowering Plants. 114.

Raphael, R. A., Taylor, E. C., and Wynberg, H. (Editors). Advances in Organic Chemistry: Methods and Results, Vol. I. 51.
Rogard, Y. General Dynamics of Vibrations. 167.
Rudnick, Dorothea (Editor). Developing Cell Systems and Their Control. 169.
Russell, R. D. and Farquhar, R. M. Lead Isotopes in Geology. 51.

RYSHKEWITCH, E. Oxide Ceramics, Physical Chemistry and Technology. 234.

SANDERSON, R. T. Chemical Periodicity. 168.
SLATER, N. B. Theory of Unimolecular Reactions. 112.
SMALES, A. A. and WAGER, L. R. (Editors). Methods in Geochemistry. 169.
STEWARD, F. C. (Editor). Plant Physiology, Vol. Ia, Cellular Organisation and Respiration. 114.
SYNGE, J. L. Relativity, the General Theory. 111.

Transactions of the Newcomen Society for the Study of the History of Engineering and Technology, Vol. XXX (1955–56 and 1956–57). 170.

UBBELOHDE, A. R. and LEWIS, F. A. Graphite and its Crystal Compounds. 112.

UNGNADE, H. E. (Editor). Organic Electronic Spectral Data, Vol. II (1952–55). 233.

VAN BUEREN, H. G. Imperfections in Crystals. 168.
VANCINI, C. A. La Sintesi dell'Ammoniaca. 232.

WILSON, J. G. and WOUTHUYSEN, S. A. (Editors). Progress in Elementary Particle and Cosmic Ray Physics, Vol. V. 111.
WITHROW, R. B. (Editor). Photoperiodism in Plants and Animals. 114.
WOLSTENHOLME, G. E. W. and O'CONNOR, C. M. (Editors). Ciba Foundation Symposium on Quinones in Electron Transport. 233.

ZIMAN, J. M. Electrons and Phonons: The Theory of Transport Phenomena in Solids. 50.

SUBJECT INDEX

The articles are arranged, with regard to subjects, in numerical order according to the Universal Decimal Classification (Brit. Std. 1000). An alphabetical key is to be found on page iv.

06 SCIENTIFIC INSTITUTIONS

061.22(100):5:070.432 An international council of sciences (Editorial). 175.
061.22(43):5:070.432 Jubilee of the *Max-Planck-Gesellschaft* (Editorial). 59.

5 GENERAL SCIENCE

5:070.432:061.22(100) An international council of sciences (Editorial). 175.
5:070.432:061.22(43) Jubilee of the *Max-Planck-Gesellschaft* (Editorial). 59.
5.001.891:070.432 Investing in research (Editorial). 119.

52 ASTROPHYSICS. ASTRONOMY

525:551.51 629.19 Space research. MASSEY, Sir HARRIE. 68.

53 PHYSICS

530.12:531.5 Gravitational waves. BONDI, H. 121.
531.5:530.12 *Idem.*
533.5:621.52 Ultra-high vacuum. LEWIN, G. 85.
535.7:612.843.3 Light quanta and vision. PIRENNE, M. H. 197.
539.124 Positronium. HEYMANN, F. F. 225.

54 CHEMISTRY

541.653 Optical rotatory dispersion. DJERASSI, CARL. 138.
541.71 Electron-deficient compounds. STONE, F. G. A. 61.
543.253 Polarography. BARKER, G. C. and MILNER, G. W. C. 26.
543.426:576.8.097.5 Fluorescent protein tracing and the fluorescent antibody method. NAIRN, R. C. 78.
546.02:621.039.32 Isotope separation by reversible chemical processes. GLUECKAUF, E. 42.
546.18:93:661.631(091) The beginnings of phosphorus manufacture. 40.
546.45:669.725 Beryllium. MCQUILLAN, M. K. and FARTHING, T. W. 11.
547.458.8:668.81 Reactive dyes for cellulose. RATTEE, I. D. 154.
547.473.2:577.15 The organic and biological chemistry of mevalonic acid. WAGNER, A. F. and FOLKERS, K. 177.
547.78:577.18:615.779 New penicillins, cephalosporin C, and penicillinase. ABRAHAM, E. P. and NEWTON, G. G. F. 92.

55 GEOPHYSICS. GEOLOGY

550.311:551.241 The Mohole. BULLARD, Sir EDWARD. 188.
551.241:550.311 *Idem.*
551.345:581.524.441:553.97 Surface patterns in Boreal peatland. SJÖRS, H. 217.
551.351.5:593.6:591.134 Problems of growth and calcium deposition in reef corals. GOREAU, T. 32.
551.46:070.432 The ocean depths (Editorial). 3.
551.51:629.19:525 Space research. MASSEY, Sir HARRIE. 68.
551.593.51 The green flash and kindred phenomena. O'CONNELL. D. J. K. 131.
552.3:92 Sir James Hall, Bt. (1761–1832). EYLES, V. A. 210.
553.97:551.345:581.524.441 Surface patterns in Boreal peatland. SJÖRS, H. 217.

57 BIOLOGICAL SCIENCES

576.312:577.15 The nucleolus of the cell nucleus. SIRLIN, J. L. 146.
576.8.097.5:543.426 Fluorescent protein tracing and the fluorescent antibody method. NAIRN, R. C. 78.
577.15:547.473.2 The organic and biological chemistry of mevalonic acid. WAGNER, A. F. and FOLKERS, K. 177.
577.15:576.312 The nucleolus of the cell nucleus. SIRLIN, J. L. 146.
577.18:615.779:547.78 New penicillins, cephalosporin C, and penicillinase. ABRAHAM, E. P. and NEWTON, G. G. F. 92.

58 BOTANY

581.13 The transport of organic substances in plants. KURSANOV, A. L. 19.
581.524.441:553.97:551.345 Surface patterns in Boreal peatland. SJÖRS, H. 217.

59 ZOOLOGY

591.134:551.351.5:593.6 Problems of growth and calcium deposition in reef corals. GOREAU, T. 32.
591.185.1:599.4 Echolocation by bats. PYE, J. D. 101.
591.552:595.799 The efficiency of a honeybee community. BUTLER, C. G. 5.
593.6:591.134:551.351.5 Problems of growth and calcium deposition in reef corals. GOREAU, T. 32.
595.799:591.552 The efficiency of a honeybee community. BUTLER, C. G. 5.
599.4:591.185.1 Echolocation by bats. PYE, J. D. 101.

61 MEDICAL SCIENCE

612.843.3:535.7	Light quanta and vision. PIRENNE, M. H. 197.
615.779:547.78:577.18	New penicillins, cephalosporin C, and penicillinase. ABRAHAM, E. P. and NEWTON, G. G. F. 92.

62 ENGINEERING. TECHNOLOGY

621.039.32:546.02	Isotope separation by reversible chemical processes. GLUECKAUF, E. 42.
621.52:533.5	Ultra-high vacuum. LEWIN, G. 85.
623.454.7(091):93	Greek fire. PARTINGTON, J. R. 162.
629.19:525:551.51	Space research. MASSEY, Sir HARRIE. 68.

66/69 CHEMICAL INDUSTRY. METALLURGY

661.631(091):546.18:93	The beginnings of phosphorus manufacture. 40.
668.81:547.458.8	Reactive dyestuffs for cellulose. RATTEE, I. D. 154.
669.725:546.45	Beryllium. MCQUILLAN, M. K. and FARTHING, T. W. 11.

9 BIOGRAPHY. HISTORY OF SCIENCE

92:552.3	Sir James Hall, Bt. (1761–1832). EYLES, V. A. 210.
93:623.454.7(091)	Greek fire. PARTINGTON, J. R. 162.
93:661.631(091):546.18	The beginnings of phosphorus manufacture. 40.

ALPHABETICAL KEY

Antibiotics 615.779
Antibodies 576.8.097
Astronomy 52

Bats, echolocation 599.4
Beryllium 546.45
Biochemistry 577.1
Biography 92
Biosynthesis 577.15
Boreal peatlands 581.524.441
Botany 58

Cell nucleus 576.31
Cellulose dyeing 547.458.8
Cephalosporin C 547.78
Chemical industry 66
Chemical separation of isotopes 546.02
Chemistry 54
Coral reef formation 551.351
Cytology 576.3

Dyes 668.8

Earth's crust 551.241
Echolocation by bats 591.185.1
Ecology
  Animal 591.552
  Plant 581.52

Electron-deficient compounds 541.71
Eye, sensitivity of 612.843.3

Fluorescence analysis 543.426

Geology 552/553
Geophysics 551.5
Gravitational waves 531.5
Greek fire 623.457
Green flash 551.593.51

Hall, Sir James 92
Hearing
  Bats 591.185
Histology 576.72
History of science 93
Honeybees 595.799

Igneous rocks 552.3
Immunology 576.8.097
International Council of Scientific Unions 061.22
Isotope separation 621.039.32

Max-Planck-Gesellschaft 061.22

Medical science 61
Metallurgy
  Beryllium 669.725
Meteorology 551.5
Mevalonic acid 547.473.2
Military engineering 623
Mohole 550.311

Nuclear physics 539.12
Nuclear technology 621.039
Nucleolus 576.312

Oceanography 551.46
Optical rotatory dispersion 541.653
Optics 535
Organic chemistry 547

Peat 553.97
Penicillins 547.78
Permafrost 551.345
Phosphorus 546.18
Physics 53
Physiological optics 535.7
Physiology
  Animal 591.1
  Human 612
  Plant 581.1

Plants
  Transport of organic substances 581.13
Polarography 543.253
Positronium 539.124
Protein tracing 576.8.097.5

Quantum theory and vision 535.7

Reactive dyestuffs 668.81
Reef corals 593.6
Relativity theory 530.12

Scientific research 5.001.891
Space
  Research 525
  Technology 629.19
Stereochemistry 541.6

Terrestrial formations 551.3

Ultra-high vacuum 533.5

Vacuum technology 621.52
Vision 612.84

Zoology 59

ऋते ज्ञानान्न मुक्तिः
पुस्तक सं. ....
आगत सं. ....
दिनांक...... ......
[illegible]

Made and printed in Great Britain at THE KYNOCH PRESS, *Witton, Birmingham,* 6, *for the publishers*
IMPERIAL CHEMICAL INDUSTRIES LIMITED, *London, S.W.1*

पुस्तकालय

The ocean depths

Science is still commonly portrayed as an entirely objective study in which one advance forms the logical prelude to the next, and new discoveries in one field are quickly and fully applied for the advancement of others. It need hardly be said here that such a picture is far removed from reality: it is a commonplace that while some fields are intensively active others are strangely neglected. One such neglected field at the moment is oceanography: it is a curious paradox that although much is known of the remotest and most inhospitable lands of the earth and we are clearly on the very threshold of exploring space, we still know remarkably little of the dark ocean depths that literally lap our modern civilization. It has been said, with a good deal of justification, that we know more about the surface of the Moon than we do of the bottom of the sea. The capture of living Coelacanths, supposedly extinct for some fifty million years, is only one indication, though a particularly striking one, of the limits of our knowledge.

To say that surprisingly little has been done in oceanography, having regard to the intrinsic interest and importance of the subject, is not by any means to belittle what has already been achieved. Some of the world's most memorable voyages have been carried out wholly or partly for the advancement of oceanography. Familiar early examples are those of Edmond Halley, in the late seventeenth century, who at great hazard penetrated as far as latitude 52½° S, and of James Cook, nearly a century later, whose epic voyages of scientific exploration twice took him round the world. Much later, another great British expedition, that of the *Challenger* (1872–76), is generally regarded as having laid the foundations of modern oceanography: in the course of this some 70 000 miles were covered—still a record for a voyage of this kind—and the results fill fifty massive volumes that remain standard works of reference. More recently, the *Discovery* committee's vessels made great contributions to our knowledge of the Southern ocean. But oceanography has long been a truly international science, and all the great maritime nations have sponsored important expeditions. Among them, to name only a very few, are Nansen's remarkable Arctic voyage in the *Fram* (1893–96); the German *National* expedition of 1889, memorable for Hensen's classic researches on the distribution of the plankton; the *Belgica* (1897–99), the first to winter in the southern ice; the Carlsberg Foundation's *Dana* expedition (1928–30); the extensive magnetic surveys of the American ship *Carnegie*; and the Pacific surveys carried out by the Scripps Institute of Oceanography. Such voyages have been an important source of new knowledge, but by no means the only one: fishery research vessels, oceanographic institutes, marine biological stations, and surveying and cable-laying ships have all made, and are still making, their contributions.

That so little is still known, despite all these agencies, is partly a measure of the immensity of the problem. More than half the surface of the earth is covered by sea more than a thousand fathoms deep, and millions of square miles of ocean are three times deeper. Impressive though they are, these figures are less daunting than they appear at first sight. The progress of oceanography does not depend on exhaustively surveying these vast regions both horizontally and vertically, any more than the progress of geology depends on surveying every square mile of the earth's land surface. Important principles can certainly be established by concentrating effort on limited objectives: a major problem today is to determine the most rewarding fields of study. Moreover, the progress of other branches of science has immensely increased the speed with which certain kinds of measurement can be made and evaluated. To quote a single example: the echo-sounder makes it possible for a continuous recording of depth to be made from a moving vessel—an immense advantage over the making of individual measurements in deep water by means of line and sinker. Again, much progress has been made in taking core-samples of deep-ocean deposits, and there are now shipborne gravimeters and towed magnetometers.

But the unevenness of progress in science as a whole is particularly manifest in oceanography. Progress certainly depends on more observational and experimental work at sea, of a carefully planned kind, but it is no longer true that 'only those who brave its dangers Comprehend its mystery'; much basic work on marine physics can best be done in laboratories ashore. Indeed, preoccupation with major economic problems—such as those of the fishing industry—certainly seems to have led to a too direct approach in some branches of oceanography. There are now welcome indications that analytical, as well as observational, methods are receiving increasing attention and that observations are being made to test theories. Much progress has been made, for example, in the theory of the generation, transmission, and dissipation of waves. Such knowledge, properly applied, can be of great practical importance to both coastal engineers and ship designers, whose approach has hitherto been, of necessity, largely empirical. What is important is to maintain a proper balance. There is a risk that the need to spend large sums on operating an oceanographic ship at sea will lead to inadequate provision for working up the results. Work in ships is an essential feature of oceanography, but the subject involves more than this; a rational attack on its problems requires the interest of scientists of many kinds and the provision of laboratories on shore and of a great variety of equipment.

When appropriate, the laboratory approach to oceanographic problems is attractive in several respects. Firstly, important results can be expected from a relatively small expenditure. Secondly, there is no lack of possible applications of new knowledge of fundamental principles. So far as shipbuilding is concerned, it is already clear that theory can make important contributions to ensuring that hull design is appropriate to the wave conditions most prevalent on the vessel's intended run. The possibilities for fishery research are sufficiently indicated by the fact that although the seas cover 72 per cent of the globe they yield only 1 per cent of our food. Again, the meteorologist and the disposer of radioactive and other industrial waste are alike interested in the factors that govern deep-ocean currents. In the last connection, it is disturbing to learn that a flow of nearly one knot has been recorded for several days at a depth of 4000 m in the Atlantic; clearly there are dangers in assuming that material dumped even in great depths will stay there indefinitely. Thirty years ago the marine biologist H. Bigelow stressed that 'the biologist . . . must perforce take as keen an interest as do his physical confrères in the modern application of mathematics to ocean dynamics'. Another important consideration has already been touched on: more attention to basic problems may well indicate how the research vessel can be used to better advantage.

Physical oceanography has perhaps already made its greatest contribution to navigation at sea, in the sense that the increasing size and power of ships makes them better able to meet severe conditions. In this field we are unlikely to see again anything as far-reaching in effect as Franklin's chart of the Gulf Stream, by use of which the eastward journey across the Atlantic could be shortened by as much as two weeks, or Maury's famous wind and current charts of 1847–60. But Admiral Smyth, a founder of the Royal Geographical Society, was over-optimistic when, in 1854, he claimed that the combination of science and practical seamanship had already nearly reduced the elements to subjection; better understanding of winds and currents, tides and waves, can certainly still help the mariner.

Oceanography has always been characterized by fruitful international co-operation. Franklin's edict that, despite the War of Independence, Cook was not to be molested by American warships during his voyages of scientific exploration is a particularly striking reminder of this tradition. Today many international agencies exist to promote oceanographic research. They include the International Council for the Exploration of the Sea, the International Association of Physical Oceanography, the International Council of Scientific Unions, UNESCO, the International Atomic Energy Agency, the World Meteorological Organization, and the Food and Agricultural Organization of the United Nations. Such a multiplicity of agencies is welcome evidence of recognition of the implications of oceanographic research, and it is indisputable that for extensive oceanic surveys—such as are currently planned for the Indian Ocean—some means of widely sharing the cost and effort is essential. But there is a danger that preoccupation with big international projects, prompted at least partly by immediate practical needs, will be to the detriment of fundamental research by individuals and small national research groups, from which important results may be expected for comparatively modest expenditure.

It has been said that activity and progress in a subject are in direct proportion to the interest that the universities take in it; by this criterion the outlook for oceanography is bleak indeed. Thus Britain, with a maritime tradition second to none, has only a single professorship of oceanography; while some other countries may be better endowed, it is certainly true to say that professorships in oceanography are few and far between. Expenditure on fundamental research in oceanography is small in comparison with the practical issues at stake. It is very much to be hoped that as more funds become available the needs of the individual will be remembered as well as those of the big international expeditions.

The efficiency of a honeybee community

C. G. BUTLER

A community of honeybees is an integrated group, the members of which can distinguish other members from intruders. Individual bees can pass on detailed information about the nature and location of food sources; they will produce a new queen only when the old one is no longer satisfactory. These different functions are discussed and illustrated in this article; the discussion of queen rearing is made particularly appropriate by the recent discovery of the chemical nature of queen substance. The photographs are by Treat Davidson.

A few years ago G. D. Glynne-Jones [1] released about two hundred worker honeybees, taken at random from a large colony, in a small greenhouse. Within a couple of hours these bees formed a cluster in one corner of the roof, and a few of them began to feed at a dish of sugar syrup on one of the benches. These bees were each given a distinctive paint mark; it was found that only about a dozen bees were leaving the cluster to collect syrup, and that they carried it back to the cluster and shared it with their companions before returning to the dish for more. Only these few bees left the cluster and collected food during the next three days. Even though they had neither queen, comb, nor brood, the bees had formed a rudimentary colony, and the fact that only a few individuals collected food for the whole group indicated the establishment of a division of labour among its members. This division between household bees concerned, for example, with comb building (figure 10), and feeding the larvae and the queen (figures 3, 11, 12), and field bees who regularly leave the nest in search of food (figures 6–9), is well established in every honeybee community. Unfortunately the way in which this major division of labour between the worker bees is effected is still unknown.

Until recently, the simple system suggested thirty years ago by G. A. Rösch [2] seemed to give a reasonable explanation of the facts. He had reached the conclusion that every worker bee normally carries out each of a number of tasks in a regular sequence between birth and death, starting as a nurse bee feeding the older larvae, and ending as a forager. He supposed that the particular job of a worker at any given time depended on her age; the degree of activity of various glands was related to this, as their secretions were necessary for different tasks. The system would need some degree of flexibility to suit the special requirements of the colony at any given time: Rösch supposed that the length of time a bee spends on one kind of job varies.

This simple theory was shown to be inaccurate in 1952, when M. Lindauer [3] published the results of very detailed observations in which he had followed the activities of several individually marked workers throughout most of their lives. He found that there is no regular sequence of hive duties, and that a household bee does several different kinds of job every day. Indeed, he found that a worker bee's life can be divided into only three distinct periods, instead of the seven or eight that had previously been proposed. In the first two or three days of life, the only task a worker appears to have is to clean out cells before the queen lays in them again; during the second period, which usually lasts about two weeks, she wanders round the nest doing various jobs; and, in the last period of her life, from when she is about eighteen days old until she dies, when she is about four weeks old, she is concerned with nest defence and foraging. So, as J. B. Free and Y. Spencer-Booth [4] have also recently shown, a bee in summer normally spends the first eighteen days of her life working in the nest, followed by a further week or ten days as a field bee.

According to Lindauer, about two-thirds of a household bee's time is spent either in resting or in patrolling the combs; he suggested that the significance of patrolling, which had been overlooked previously, might lie in the fact that it would enable each worker to find out what different jobs required attention, and that response to simple stimuli would result in each bee doing any work required. In support of this theory, Lindauer instanced examples of individual workers appearing to act quite independently of one another. For instance, when a larva is full-grown, a capping of wax is used to close its cell, but this is not made by a few bees only. On one occasion 657 visits, lasting nearly 6½ hours altogether, were

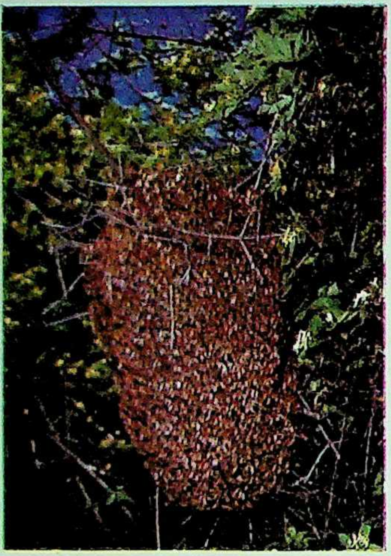


FIGURE 1 – *Swarm.*

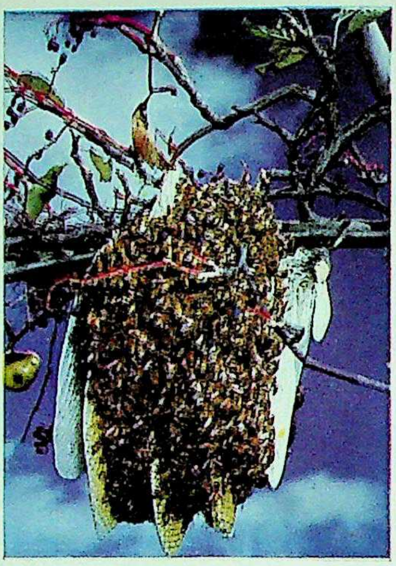


FIGURE 2 – *Colony with nest in the open. Honeybees usually nest in dark cavities, such as hollows in trees.*

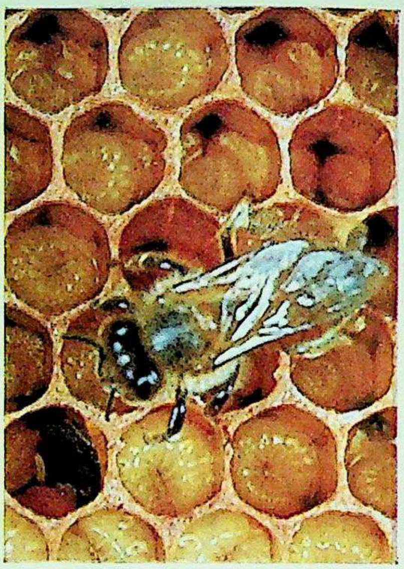


FIGURE 3 – *Nurse bee on brood comb, showing larvae and cell with egg.* (×3 *approx.*)

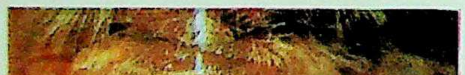
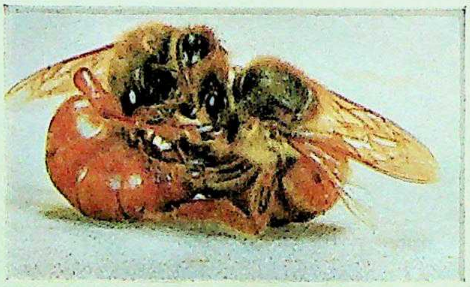


FIGURE 4 – *Underside of a worker's abdomen, showing wax scales.* (×11 *approx.*)

FIGURE 5 (*right*) – *Queens fighting.* (×2¼ *approx.*)



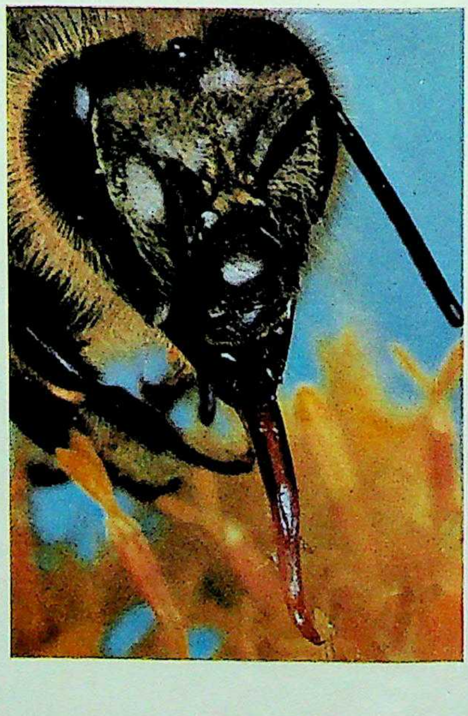


FIGURE 6 – *Tongue of nectar-seeking worker.* (×11 *approx.*)

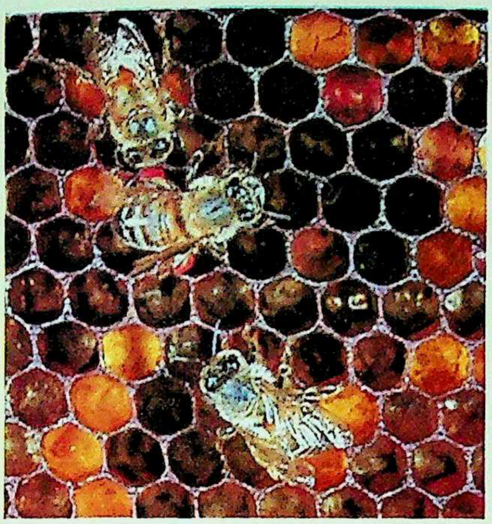


FIGURE 7 – *Workers on pollen-storage cells.* (× 1½ *approx.*)

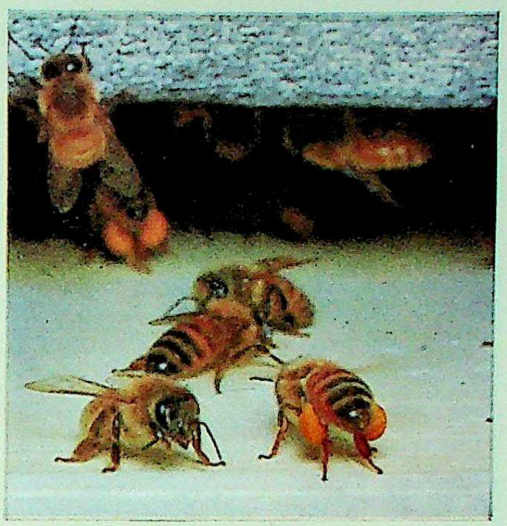
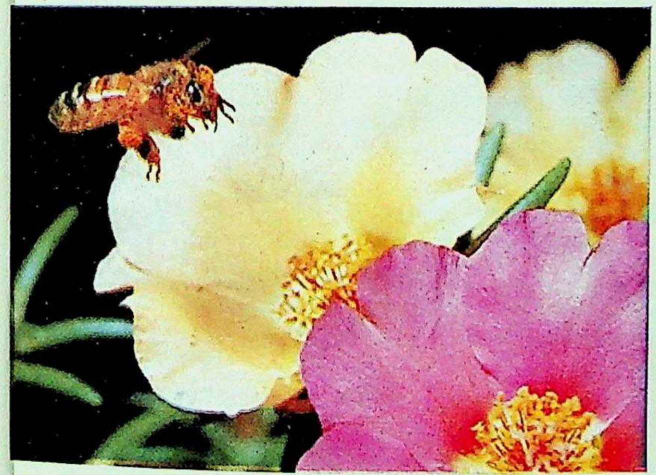


FIGURE 8 – *Forager* (*with pollen*) *and guards at hive entrance.* (× 2 *approx.*)



GURE 9 – *Forager approaching a flower, showing clumps of pollen body and pollen-loads on hind-legs.* (× 2 *approx.*)

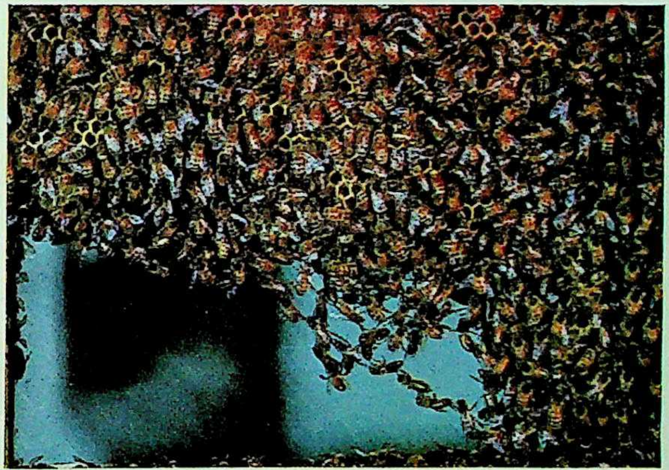
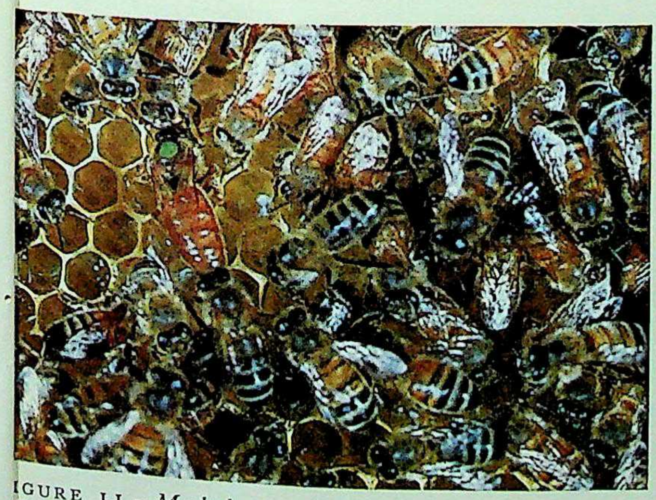


FIGURE 10 – *Comb building.*



IGURE 11 – *Marked queen* (*with a green spot*) *with 'court'. he worker near her tail is probably collecting queen substance, the e near her head is feeding her.* (*Natural size.*)

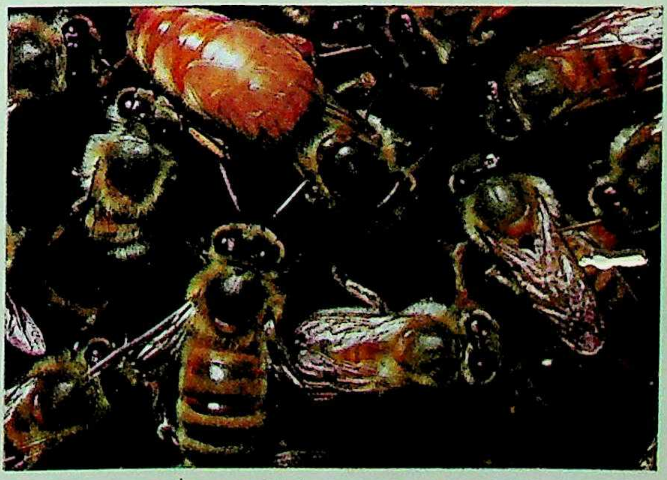


FIGURE 12 – *Queen with 'court', showing workers examining her with their antennae.* (× 2½ *approx.*)

made by different bees before the capping of one cell was completed. Not only does such a job proceed in a haphazard manner, but sometimes a bee will remove a particle of wax that another bee has just laboriously added and use it in the capping of a neighbouring cell. A similar lack of co-operation is often seen among a group of guard bees dragging an intruder out of their nest. Each will often pull in a different direction [5].

It seems probable that Lindauer is correct, and that each bee reacts independently to the stimuli she receives from young and old larvae, intruding bees, half-capped cells, and so on. But it also seems probable that a bee spends more time doing those jobs for which she is best adapted at the moment. For instance, a bee whose wax glands are secreting wax abundantly (figure 4) may spend a lot of her time building combs.

It is difficult to see how the important division between household and field bees is effected. Perhaps, as C. R. Ribbands has suggested [6], converting nectar into honey, and comb building, done mainly by the older household bees, may be associated with an increasing tendency to leave the vicinity of the broodnest, this tendency eventually culminating in foraging.

Ribbands [7] showed that anaesthetization of young worker-bees with carbon dioxide causes them to begin foraging at an abnormally early age. O. Mackensen [8] found that queens that are prevented from mating seldom start laying eggs until they are about three weeks old, whereas similar queens, anaesthetized with carbon dioxide when about a week old, begin laying a few days later—at the same age as a queen that has mated normally. It seemed possible, therefore, that anaesthetization with carbon dioxide in some way aged the bees; however, Ribbands [7] found that this was not so, although it reduced the expectation of life of a worker by making her start the hazardous occupation of foraging at an abnormally early age. It may be that anaesthetization with carbon dioxide has a sudden effect on a bee's metabolism similar to one that perhaps occurs suddenly when a queen mates and starts egg laying, or more gradually when a worker forsakes household duties in favour of foraging.

It has already been mentioned that the average life of a worker bee in summer is about four weeks. However, those that are reared in autumn live much longer [4]. During early winter, fat and protein accumulate in a worker's fat-body, and A. Maurizio showed [9] that the long winter life appears to depend to some extent on this accumulation. Long-lived, 'winter', bees can be produced in summer in queenless colonies; the queenless bees accumulate fat and protein reserves.

Only bees reared in autumn survive into winter, and a critical period in a colony's life occurs in spring, when the old bees that have survived the winter are dying off rapidly and new ones are being reared. The death of many colonies early in some years may well have been caused by protein deficiency, because too little pollen was stored in the nests during the previous summer; the old bees probably died before enough young bees were reared to make the colonies viable.

Although the system of division of labour, in which a household bee wanders around the nest doing any job she finds, seems rather haphazard, it appears to be both efficient and sufficiently adaptable to allow even for sudden, extreme changes in a colony's requirements. When, for example, an experimental colony is made up entirely of foragers, some of them revert to brood-rearing and similar household tasks [10]; conversely, when one is made up entirely of young household bees, some begin foraging at an unusually early age [11, 12].

On leaving her nest for the first time, the average forager does not wander from one kind of flower to another searching for food. In fact very few bees ever leave the nest before being stimulated by the dance-language of successful foragers [3]. From this they learn that a certain kind of food, associated with a particular scent, is at a given distance and direction from the hive [13–17]. A small proportion of a colony's foragers leaves the nest without any knowledge of where there is food; these bees scout for the whole colony and are responsible for finding new sources as they occur. Unlike most foragers, who settle down to exploit a source thoroughly, scout bees tend only to visit a particular source a few times before they start looking for another [18]. When they are searching for new sources of food, scout bees are attracted by small, brightly coloured—especially blue and yellow—objects, such as flowers or even discs of paper. They seldom settle on them, however, unless the objects have a scent [19]. When a scout bee is attracted to a flower by its colour she hovers over it and, as soon as she perceives its scent, settles on it and extends her tongue to search for nectar in any narrow crevices (figure 6). If she finds plenty of food she collects a load before returning to her nest, where, by means of the 'dance-language' [13–17], she informs other bees of its whereabouts and recruits some of them to exploit the source.

On finding a rich source of food, a bee often exposes her scent-producing gland. Scent dispersed in this way helps searching foragers to find the food, and there is evidence that, although the scent is attractive to all bees, it is particularly attractive to other members of the scent-dispersing bee's colony [20, 21]. H. Kalmus and C. R. Ribbands [21] demonstrated that such differential attraction is a consequence of odour differences, and that every colony of honeybees possesses its own particular odour, shared by all its members and differing from that of other colonies. They found that the odour of a colony depends on the kind of food its members have been sharing among themselves. Different colonies, even in the same apiary, collect different proportions of their food from the various kinds of flowers available at any given time. If this explanation of the origin of distinctive colony-odours is correct, it follows that a number of colonies compelled to share a single source of food should all acquire the same odour. Ribbands [6] obtained striking confirmation of this by removing the stores of food from a number of colonies and taking the colonies to a heather moor where the only source of food was *Calluna vulgaris*. At the end of two months all the colonies had, apparently, a common odour, as their members could be freely interchanged.

Colony odour plays an important role in enabling bees that are guarding the entrance of their nest against intruders to distinguish between their own sisters and bees from other colonies [22, 23]. But although the workers of a colony share a common odour, the queen does not appear to share the same odour as the workers; these seem to be unable to distinguish between their own mated or virgin queen and similar queens from other colonies introduced into their nests; if they can make the distinction, they ignore such differences [24]. Nevertheless, worker bees certainly distinguish between mated and virgin queens by their different odours [25], and are attracted towards a queen, even one that has been dead for several months, by her body odour [26].

Queen odour also probably helps a queen to find a rival that has appeared in her nest and so gives her an opportunity to destroy her (figure 5), unless the workers prevent the two queens from meeting. It also seems possible that the worker bees in a swarm (figure 1) are aware that their queen has swarmed with them by an odour from her body. Certainly if she fails to leave the nest with them or to join them when they cluster, the workers soon notice her absence and return to their old nest.

When worker bees lose their queen they try either to rear a new one or, sometimes, to join a group of bees that has one. Generally speaking, the presence of a laying queen in a colony suffices to inhibit both development of the workers' ovaries and queen rearing by them. Such inhibition is not brought about by the bees realizing that their queen is in the nest because of her pervading odour. In fact the sight, sound, and scent of a queen do not cause such inhibition, which only occurs when some members of the colony make frequent contact with their queen [27–30]. This led the author [29] to suggest that some of the members of a honeybee colony obtain something—queen substance—from their queen that they share with other members of their colony and that, when each bee receives enough, inhibits development of the ovaries and queen rearing by the bees. Experiments in which small groups of workers were allowed access only to limited parts of the bodies of their queens for two or three days but remained inhibited from queen rearing indicated that this substance is present all over a queen's body [29]. This queen substance could also be recovered from the crops of worker bees caught while licking their queen (figure 11); it is almost certainly shared among all the members of a colony in regurgitated food [31].

Many experiments indicated a quantitative relationship between the amount of queen substance and the inhibition of queen rearing. However, although much circumstantial evidence in favour of the existence of queen substance accumulated, it was not until 1957, when a quick and fairly sensitive method of bioassay was developed, that it became possible to show that an extract in ethanol of a substance from a mated queen's body, when fed to queenless worker bees, inhibits them from building queen cells [32].

Having established the existence of queen substance, one of the next problems was to discover its source. The fact that the mandibular glands of a queen are proportionally very much larger than those of a worker led C. G. Butler and J. Simpson [33] to seek and find queen substance in them, and to suggest that the queen distributes it over her body during grooming.

Under natural conditions new queens are generally reared to replace old ones either by supersedure or swarming. (Supersedure differs from swarming in that the colony replaces its queen with a young one without forming a new colony.) Many theories have been suggested to explain both supersedure and swarming: there is now no doubt that in connection with both, queen rearing happens because the workers receive too little queen substance [35]. With swarming, this seems to be brought about in two ways. Firstly, when a colony is overcrowded, the workers get too little queen substance, even though the queen produces plenty in her mandibular glands, either because the bees who normally lick it from the queen's body are unable to do so properly in the crowded conditions, and are unable to distribute what they do collect widely enough, or, possibly, because the queen is jostled around so much that she is hindered in grooming, and hence in distributing queen substance over her body. Secondly, when a colony that is not overcrowded builds queen cells preparatory to swarming, its queen is actually producing too little queen substance. Both queens that have been superseded and those that have accompanied swarms from uncrowded colonies contain less than one-quarter the amount of queen substance of an average, mated, laying, queen, whereas queens of swarms from overcrowded colonies have just as much [35].

Queen substance has now been isolated, by R. K. Callow, and N. C. Johnston, as waxy crystals, 0·13 μg of which will inhibit a worker from queen rearing for at least twenty-four hours [36]. This crystalline material has been identified by these workers as 9-oxodec-2-enoic acid,

$$CH_3.CO[CH_2]_5.CH{:}CH.COOH,$$

synthesized and shown to be biologically active [37].

To sum up, the efficiency of a honeybee community (figure 2) owes much to (*a*) a division of labour, not only between the queen and the thousands of workers, but also among the workers themselves; (*b*) the attraction workers have for each other and the attraction of a queen for workers—which keep the colony together; (*c*) a system of communication between the workers by means of which detailed information is given about food sources; (*d*) a method by which workers can distinguish between members of their own colony and intruders from other colonies; this is essential to nest defence; (*e*) a system of control which prevents ovary development and queen rearing by workers so long as a satisfactory queen is present, but which allows the colony to replace its queen by a young one and also to reproduce itself by swarming when conditions are favourable.

REFERENCES

[1] Glynne-Jones, G. D. Personal communication.
[2] Rösch, G. A. *Z. vergl. Physiol.*, **2**, 571, 1925.
[3] Lindauer, M. *Ibid.*, **34**, 299, 1952.
[4] Free, J. B. and Spencer-Booth, Y. *Proc. R. ent. Soc. Lond., A*, **34**, 141, 1959.
[5] Butler, C. G. and Free, J. B. *Behaviour*, **4**, 263, 1952.
[6] Ribbands, C. R. 'The behaviour and social life of honeybees', Bee Research Association Ltd., London. 1953.
[7] *Idem. J. exp. Biol.*, **27**, 302, 1950.
[8] Mackensen, O. *Glean. Bee Cult.*, **79**, 273, 1951.
[9] Maurizio, A. *Bee World*, **31**, 9, 1950.
[10] Moskovltevic, V. *Ibid.*, **21**, 39, 1940.
[11] Nelson, F. C. *Amer. Bee. J.*, **67**, 242, 1927.
[12] Gontarski, H. *Umschau*, **49**, 310, 1949.
[13] von Frisch, K. *Zool. Jber. Abt.* 3, **40**, 1, 1923.
[14] *Idem. Öst. zool. Z.*, **1**, 1, 1946.
[15] *Idem. Experientia*, **2**, 397, 1946.
[16] *Idem. Ibid.*, **5**, 142, 1949.
[17] *Idem. Naturwissenschaften*, **38**, 105, 1951.
[18] Oettingen-Spielberg, T. *Z. vergl. Physiol.*, **31**, 454, 1949.
[19] Butler, C. G. *Proc. roy. Soc. Lond., B.*, **138**, 403, 1951.
[20] von Frisch, K. and Rösch, G. A. *Z. vergl. Physiol.*, **4**, 1, 1926.
[21] Kalmus, H. and Ribbands, C. R. *Proc. roy. Soc. Lond., B*, **140**, 50, 1952.
[22] Butler, C. G. and Free, J. B. *Behaviour*, **4**, 263, 1952.
[23] Ribbands, C. R. *Proc. roy. Soc. Lond., B*, **142**, 514, 1954.
[24] Butler, C. G. and Simpson, J. *Bee World*, **37**, 105, 1956.
[25] Butler, C. G. 'The World of the Honeybee', Collins, London. 1954.
[26] *Idem. Experientia*, 1960. (In press.)
[27] Huber, F. 'Nouvelles observations sur les abeilles' (second edition), 1914. Dadant, Hamilton, Illinois. 1926.
[28] Müssbichler, A. *Z. vergl. Physiol.*, **34**, 207, 1952.
[29] Butler, C. G. *Trans. R. ent. Soc. Lond.*, **105**, 11, 1954.
[30] de Groot, A. P. and Voogd, S. *Experientia*, **10**, 384, 1954.
[31] Butler, C. G. *Proc. R. ent. Soc. Lond., A*, **31**, 12, 1956.
[32] Butler, C. G. and Gibbons, D. A. *J. Insect Physiol.*, **2**, 61, 1958.
[33] Butler, C. G. and Simpson, J. *Proc. R. ent. Soc. Lond., A*, **33**, 120, 1958.
[34] Butler, C. G. *Insectes sociaux*, **4**, 211, 1957.
[35] *Idem. Proc. R. ent. Soc. Lond. A*, **35**. (In press.)
[36] Butler, C. G., Callow, R. K., and Johnston, N. C. *Nature, Lond.*, **184**, 1871, 1959.
[37] Callow, R. K. and Johnston, N. C. *Bee World*, **41**, 152, 1960.

Beryllium

M. K. McQUILLAN and T. W. FARTHING

Beryllium has been known since the end of the eighteenth century and has for some time been used in copper alloys, but it is the development of atomic reactors that has made the preparation and working of the metal an important industrial process. The very low neutron cross section of the metal makes it valuable as a canning material for the fuel rods in gas-cooled reactors. The processing of the metal presents many unusual features, resulting from both the inherent physical properties of the metal and its high toxicity.

Beryl, the mineral which gives beryllium its name and which is still the only commercial source of the metal, has been known for thousands of years, both as a semi-precious gemstone and, in its green single-crystal form, as emerald. The history of the metal itself, however, is much shorter. It was isolated in 1798 by Vauquelin, but its debut as an industrial material occurred only about thirty years ago, when it was observed that small additions of beryllium to copper could give the latter the strength and elasticity which pure copper notoriously lacks but without destroying the high thermal and electrical conductivity upon which so many of the uses of copper depend.

Copper-beryllium alloys still account for about three-quarters of all beryllium extracted from its ores, but in the last fifteen years or so the pure metal has become of interest, because of its exceptional combination of properties, as a material for use in nuclear reactors. Without this development it is very probable that pure beryllium would still be little more than a metallurgical curiosity, although, as we shall see later, it has other possible applications.

From the point of view of a nuclear engineer, the most important single property of beryllium metal is its very low absorption of thermal neutrons—the lowest of any structural metal. Neutrons are the means by which nuclear reactions are propagated in reactors, and reactor efficiency thus depends on keeping adventitious neutron absorption and losses to a minimum. Neutron-absorption characteristics are therefore among the principal factors determining choice of reactor materials, and hence beryllium is naturally favoured in this respect. It has, moreover, a very low atomic weight, which fits it for use as the moderator, the material surrounding the fuel and serving to slow down the neutrons produced in the fission process so that they can more efficiently interact with further uranium nuclei.

Although the essential characteristic of beryllium, from the nuclear engineering point of view, is its low neutron-absorption it is also of interest to consider its general behaviour under neutron bombardment, as this will largely govern its behaviour in a reactor. When beryllium is bombarded with fast neutrons (about 2 MeV) the main nuclear reactions occurring are those producing helium and tritium (^{3}H):

1. (n, 2n) $^9\text{Be} + \text{n} \longrightarrow {}^8\text{Be} + 2\text{n}$ (200 millibarns[1])

$$^8\text{Be} \longrightarrow 2\,{}^4\text{He}$$

2. (n, α) $^9\text{Be} + \text{n} \longrightarrow {}^6\text{He} + {}^4\text{He}$ (50 millibarns)

$$^6\text{He} \xrightarrow{\beta} {}^6\text{Li}$$

$$^6\text{Li} + \text{n(thermal)} \rightarrow {}^4\text{He} + {}^3\text{H} \quad \text{(950 barns)}$$

$$^3\text{H} \xrightarrow{\beta} {}^3\text{He}$$

Irradiation experiments suggest that, at temperatures below 600° C, gases produced during neutron irradiation will not have any serious effect on the properties of beryllium.

Discovery of the promising nuclear properties of beryllium led American workers to undertake, in the period 1945–50, an intensive study of its technology, on the results of which a considerable part of present-day industrial practice has been based. Since then there has been some slackening of American interest in beryllium because the general trend of reactor design for power production in America has been towards a pressurized-water reactor operating at about 300° C, for which beryllium would have inadequate corrosion resistance. Now, however, a developing interest in new types of reactor—such as the so-called package reactors, which are intended to provide a flexible and transportable power source, and high-temperature forms of the gas-cooled reactor—is again bringing beryllium to the forefront of the American nuclear-energy programme.

---

[1] The barn (10^{-24} cm^2) is the unit of neutron cross section.

In Britain, reactor design has taken a different course. The reactors favoured here for power production are graphite-moderated and gas-cooled, as typified by Calder Hall and its successors, and in such reactors beryllium can be expected to make an excellent sheathing (canning) material for the uranium fuel, as it is compatible both with the carbon dioxide coolant and with the graphite core into which the sheathed fuel is inserted. Interest in beryllium among British nuclear engineers has consequently been maintained, though the metal has not yet been used because so far operating temperatures have been such that the temperature at the surface of the fuel can has not exceeded 400–500° C, at which level more readily available materials, notably magnesium, can be used. At higher operating temperatures such as are proposed for a new type of experimental reactor, known as the Advanced Gas-cooled Reactor (AGR), the surface of the fuel can is expected to reach about 600° C, at which present canning materials will have insufficient strength and stability. They must therefore be replaced by material having greater high-temperature strength.

Higher operating temperatures will also involve a change in the nature of the fuel itself. When irradiated with neutrons, uranium swells and changes its shape. At the working temperature of Calder Hall and of the nuclear power stations now being built in Britain this is not a serious problem, but under the more exacting conditions of the AGR the swelling of the uranium would create much greater difficulties. For the AGR the fuel will, in fact, be uranium oxide, which has considerably better dimensional stability under irradiation. The proposed change in fuel, however, brings with it new problems, because uranium oxide suffers from the disadvantage of poor thermal conductivity: in order to provide sufficient cooling, the diameter of the individual fuel rods has to be appreciably decreased, thus increasing the proportion of sheathing to fuel in the core. In the new design, therefore, sheathing-material of low neutron-absorption is more important than ever. It is fortunate that in beryllium, which from this point of view is the obvious material to use, we have also a material of reasonably high melting point (1285° C as compared with 650° C for magnesium), which promises good high-temperature strength; thus two of the major problems of the AGR concerning constructional materials can be solved simultaneously. It has therefore been decided to use beryllium cans for the AGR, and in consequence industry is now for the first time dealing with the metal in quite large quantities. As we shall see, this is no easy task, for reasons connected with both metallurgy and with the health of the workers concerned.

PROPERTIES

Beryllium is a true metal, in that it has the electrical and thermal properties and the lustrous appearance by which metals are usually recognized. At the same time it has other characteristics which set it apart from the general run of metals and put difficulties in the way of its industrial use.

One of the most important of these is its behaviour on deformation. The malleability which allows metals to be worked into complex shapes is a consequence of their simple crystal-structure together with the absence of direct and specific bonds between one atom and the next. Under an applied force it is possible for atoms in one plane of a metal crystal to move with respect to those in an adjacent parallel plane[1] without disrupting the crystal, the sum of a large number of such atomic-scale movements making up a macroscopic shear. Naturally, different forces will be required to produce shear on different crystal planes. In normally malleable metals the shearing process occurs on one type of plane after another as the component of applied force, resolved in the appropriate direction of shear for each type of plane, exceeds the critical magnitude for the plane in question. The higher the degree of symmetry in the crystal lattice the greater the number of planes on which the shearing process (slip) can take place and the smaller the range of magnitude for the critical shear-stresses involved.

Beryllium has a crystal structure of the close-packed hexagonal type. Its axial ratio, however, is 1·568, which is too small for ideal close-packing; beryllium is not, therefore, among the metals having the highest degree of symmetry. Nevertheless, there is no *a priori* reason why a metal having such a crystal structure should not be fairly readily deformed by making use, for instance, of slip on the pyramidal $\{10\bar{1}1\}$ planes, as in titanium, which has a closely similar crystal structure and which can be very ductile indeed. Normal commercial beryllium, however, does not in fact deform in this way. Slip can occur on the basal (0001) plane and on the prismatic $\{10\bar{1}0\}$ planes, the slip directions being $\langle 11\bar{2}0 \rangle$. Unfortunately at room temperature the critical resolved shear-stress for slip on the $\{10\bar{1}0\}$ planes is high compared with the stress needed to bring about cleavage fracture on the (0001) plane, and consequently cracking tends to occur before very much $\{10\bar{1}0\}$ slip takes place. In practice, therefore, room-temperature deformation is confined to slip on the (0001) plane, and forces applied in the majority of directions in the crystal will cause brittle fracture.

---

[1] This process does not, of course, involve the movement of whole planes of atoms simultaneously—such a movement would require impracticably large forces—but is accomplished by the propagation through the crystal of imperfections in the lattice known as 'dislocations'.

This inconvenient behaviour on deformation means that the polycrystalline aggregate making up a normal piece of beryllium will not have sufficient ductility for engineering purposes unless special care is taken to ensure that the grain structure is such as to allow maximum advantage to be taken of the slip that can occur before fracture takes place. It is essentially a question of achieving a sufficiently small grain-size coupled with a suitable grain-orientation, and the technology of beryllium is considerably complicated by the requirements of grain-structure control.

Another characteristic of commercial beryllium that has considerably influenced the development of working procedures is the complex relationship between temperature and ductility, the form of which, as suggested by early work on beryllium [1], is illustrated in figure 1. A rather more detailed picture may be revealed by current investigations—indeed, it has already been shown by A. B. Brown, F. Morrow, and A. J. Martin [2] that suitably treated beryllium can have high ductility in the temperature region of about 600° C, where a minimum in the ductility/temperature curve is normally observed—but, at present, production practice is still governed very much by the need to work at a carefully controlled temperature in order that sufficient ductility can be reliably produced.

Beryllium has a number of other chemical and physical properties—some advantageous, some the reverse—which set it apart from the common run of metals. It is, for example, exceptionally light, having a density of only 1·85 g/cm^3, and it is certainly the lightest of potential structural materials. In beryllium, lightness is also coupled with quite high strength and—even more important—very great rigidity; its ultimate tensile stress ranges from 20 to 50 tons/in^2 at room temperature, and its modulus of elasticity (44×10^6 lb/in^2) is far above that of most other metals. The thermal conductivity of beryllium at room temperature (0·385 cal cm/cm^2 sec °C) is exceeded only by those of gold, silver, copper and aluminium, and its specific heat (0·425 cal/g °C) is also high.

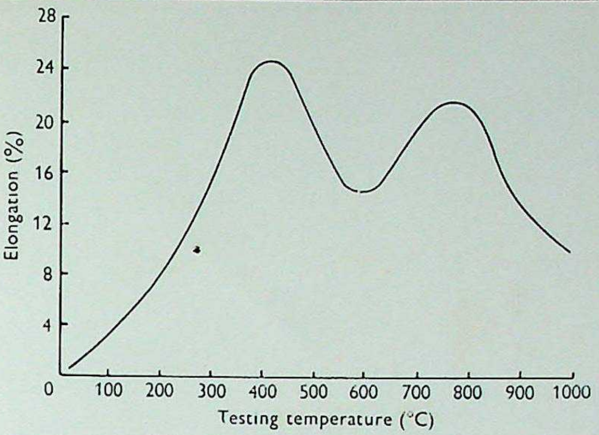


FIGURE 1 – *The effect of temperature on the ductility of polycrystalline beryllium as measured by elongation to fracture in short-time tensile test.*

Chemically, beryllium shows many similarities to aluminium. It is extremely reactive *per se*, but in air or oxidizing atmospheres at room temperature it forms a thin, adherent, impervious surface-film of inert oxide, which can protect the metal against certain types of chemical attack, giving it a spurious nobility. The protective characteristics of beryllium oxide (beryllia) are also evident in the resistance to tarnishing or oxidation shown by some beryllium-containing alloys. Silver, for instance, can be made appreciably more resistant to tarnishing by the addition of very small quantities of beryllium.

Beryllium dissolves rapidly in sulphuric and hydrochloric acids, but in nitric acid the rate of attack depends on acid concentration. Under some conditions the rate of attack by nitric acid falls to zero, the metal becoming passive. This is probably explained by the ease with which a protective coating of beryllia may be re-formed under oxidizing conditions. The corrosion of beryllium by neutral aqueous solutions is extremely variable, largely because of the important part played by very small quantities of impurities in the metal or the corroding medium, which cannot always be accurately controlled. For instance, the presence of a small amount of the ion of a more noble metal in the corrosive solution may give rise to pitting, as in the case of aluminium. As the quality of the metal available for research in the past has not been consistent, interpretation of the corrosion behaviour of beryllium is not always easy.

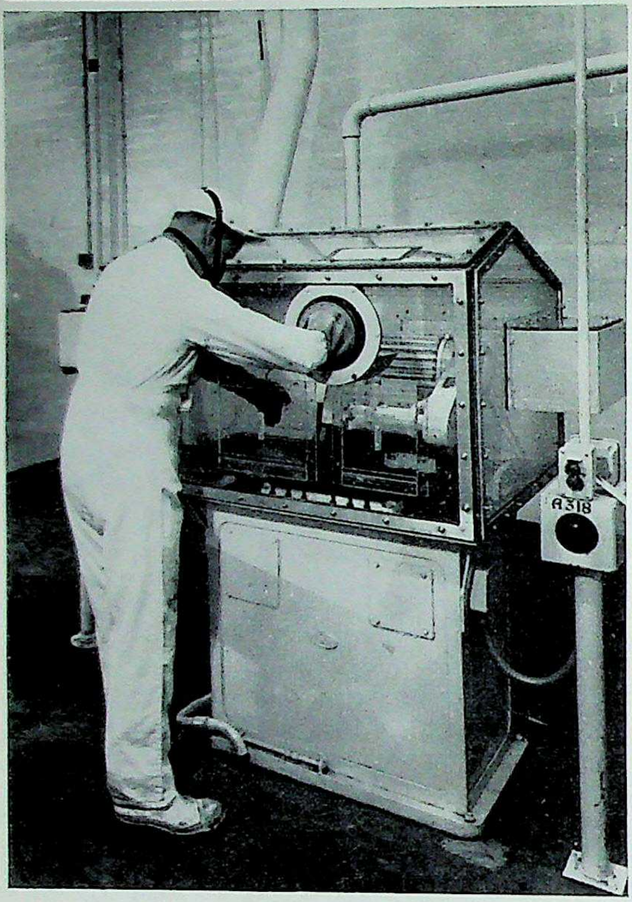


FIGURE 2 – *Components are being cut to length in a glove-box. Face masks are worn while equipment is being commissioned.*

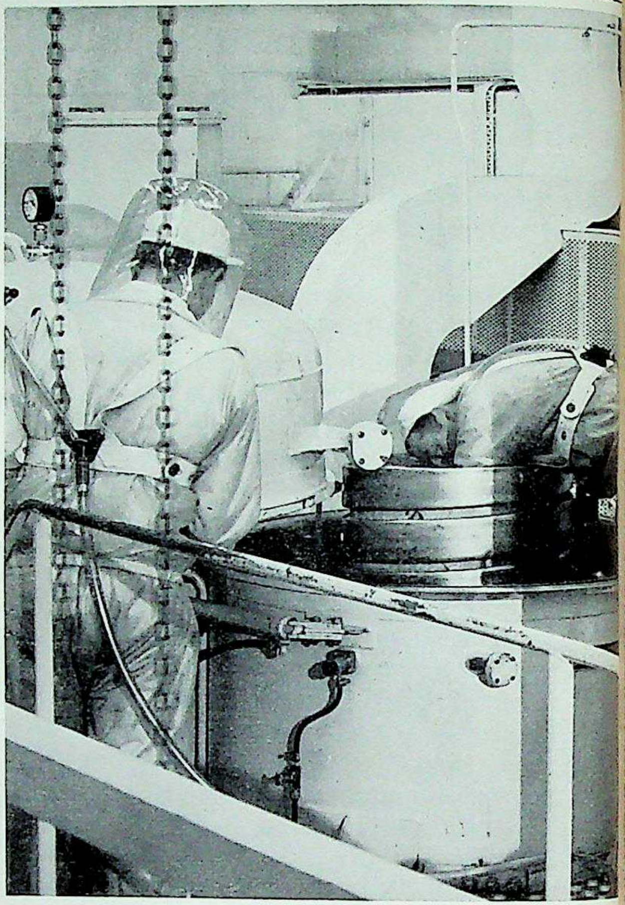


FIGURE 3 – *Maintenance staff working on a sintering furn wear special protective clothing.* (*Figures 2 and 3 by courtesy Metals Division of Imperial Chemical Industries Limited.*)

In the conditions foreseen for the AGR, that is, in carbon dioxide at 600° C, beryllium behaves satisfactorily as long as the gas is dry, but when water vapour is present the material is liable to suffer sudden marked increases in corrosion rate after a lapse of time. This phenomenon is known as 'breakaway', which graphically describes the often catastrophic effects of the change in the corrosion rate.

Above about 700–800° C, beryllium begins to react at an appreciable rate with air. For hot-working processes, therefore, some form of protection against oxidation must be provided, not only because it is difficult to work with a material which forms a heavy surface scale, but also because the dissemination of beryllium oxide in the atmosphere constitutes a severe health hazard. Molten beryllium is extremely reactive and will attack most materials with which it comes in contact; it can, however, be contained in crucibles made of beryllium oxide.

THE PROCESSING OF BERYLLIUM

Beryllium is extracted from beryl (beryllium aluminium silicate) by a process having four principal stages:

1. Separation of the beryllium content of the ore.
2. Conversion to beryllium hydroxide.
3. Production of a halide of beryllium.
4. Reduction of the halide to metal.

The separation stage involves treatment with fluoride or sulphate, producing a water-soluble beryllium salt from which the hydroxide is precipitated with alkali. The choice of the halide into which the hydroxide is then converted depends on the reduction method to be used. For reduction with magnesium, the method commonly used in the United States, the material required is beryllium fluoride, but for an alternative process, developed principally in France and involving electrolysis in a fused-salt bath, beryllium chloride is used. The two processes yield beryllium metal in different forms, magnesium reduction producing material described, from its appearance, as 'pebble', and the electrolytic method giving dendrites or flakes.

Neither reduction process gives 'raw metal' sufficiently pure for direct use; if processed without further purification the corrosion resistance of the product would be poor and very variable. It is necessary, therefore, to remove impurities—particularly residual volatile substances such as magnesium, sodium, and halides—in order to ensure a final product that has satisfactory corrosion-resistance and is uniform in its other properties. This is done by vacuum melting, usually in an induction furnace using a beryllia crucible; electric-arc melting has been used, but only on a very much smaller scale. The grain structure of the cast material is such that its room-temperature ductility is close to zero, and the room-temperature mechanical properties obtained after processing it are not as good as those of material produced from fine-grained starting stock. To achieve optimum properties, therefore, it is necessary to convert the ingot to fine powder, and consolidate it once more by one of several different techniques to produce fine-grained metal. Powdering is done first by turning the ingot to swarf and then grinding to 200 mesh. Ball mills are sometimes used, often with tungsten carbide balls to reduce iron contamination, but when contamination must be at a minimum the swarf is ground between beryllium plates in an attrition mill.

Methods available for consolidating the powder range from the old-established powder-metallurgy technique of cold pressing, followed by sintering, to the method known as pressureless sintering. This consists of pouring loose powder into a suitable die and consolidating it, without applying pressure, by heating it in vacuum at a temperature very little below the melting point. More efficient consolidation is achieved, however, by compressing the powder at elevated temperatures, thus combining the compacting and sintering operations, and though this puts very severe stress on the compacting tools and can be very time-consuming, it has become the most common procedure. The operation is done in vacuum at 1000–1100° C under an applied pressure which may be anything from about 50 lb/in^2 up to about 1 ton/in^2. The actual pressure is determined by such practical factors as the type of die material available and production economics. Hot pressing was first applied to beryllium in the United States, as an alternative to casting for producing large slabs of metals required for reflectors in an early materials-testing reactor. The method has since been developed there to a stage at which pieces of metal several inches thick and about 10 ft^2 in area can be made on a regular production basis. Cylinders about 4 or 5 ft in diameter and the same in height, and other shapes weighing up to 2000 lb, are also being successfully manufactured.

Even in the fine-grained condition, beryllium has insufficient ductility at room temperature (about 3–5 per cent elongation) to make cold working practicable, but further fabrication can be carried out by extrusion, rolling, forging, and other normal working techniques at temperatures sufficiently high to bring the prismatic $\{10\bar{1}0\}$ slip planes into operation. In practice, working temperatures range from 400° C upwards. At 400° C advantage can be taken of the first peak on the ductility/temperature curve; the maximum working temperature is usually 1000–1050° C.

During mechanical working the beryllium crystals tend to arrange themselves in such a way with respect to the direction of working that slip occurs easily in this direction. In normal wrought beryllium, which is most commonly produced by extrusion, ductilities of up to 30 per cent can be expected in the longitudinal direction. Elongations as high as 50 per cent in the rolling direction have been obtained in specially prepared sheet, but in directions other than the rolling direction much lower ductilities are exhibited. When there are two working directions, as for example in cross-rolling, high elongations can occur in both directions, but the ductility in the direction perpendicular to the plane of the sheet is again very low.

THE TOXICITY OF BERYLLIUM

One of the most important points to be taken into account in processing beryllium is that it is highly toxic. This manifests itself in two main forms—an acute irritation of the respiratory tract, including chemical pneumonia, and a chronic disease of the lungs, now known as berylliosis. Like many other metals, beryllium may also cause skin reactions such as dermatitis and ulcers, particularly if it finds its way into an open wound, though compared with the effects on the lungs this is of less importance. Various views have been expressed on the relative toxicities of different beryllium compounds, but from the health point of view, all compounds must be regarded as suspect. Much still remains to be learned of berylliosis, but it is becoming clear that only relatively few people are susceptible to it: it appears to have aspects akin to those of an allergy. A very disconcerting feature of the disease is that its onset may be delayed until many years after the last time at which the victim was exposed to beryllium.

In working with the metal or its compounds, the danger arises not from the solid material, which can be handled quite safely at room temperature, but from breathing finely divided material, which might be retained in the lungs: the beryllium content of the working atmosphere must therefore be strictly controlled. Work done in the United States, where the beryllium hazard was first recognized in the early 1940s, has produced recommendations for tolerable (extremely low) atmospheric levels that are internationally accepted. Within the plant, the atmospheric concentration, averaged over an 8-hour working shift, should not exceed 2 $\mu g/m^3$. Further, no matter how low the shift average, no worker must at any time be exposed to an atmospheric concentration of more than 25 $\mu g/m^3$. In the neighbourhood of a plant the atmospheric level, averaged over a month, must be less than 0·01 $\mu g/m^3$.

In order to keep within these very stringent safety-limits, a number of precautions must be taken. The difficulty is enhanced not only by the fact that powdering is an intrinsic part of the processing, but also by the brittleness of the metal. This, aggravated by the need to work without a lubricant in order to prevent contamination, results in the formation of much finely divided material during machining. Machine tools and other equipment are therefore set up in individual enclosures, from which the air is continuously pumped (figure 2). Handling is done with rubber gauntlets let into the sides of the enclosure. Special clothing is worn by all who work inside the plant, and the workers must take a shower before resuming their own clothes. The working garments are laundered within the plant itself. Where there is considered to be any unusual risk for which the normal precautions might not be adequate, air-fed pressurized suits are worn (figure 3).

All air extracted from the plant is filtered before release to the external atmosphere—through a 150 ft stack—by a bank of 'absolute' filters. Within the plant, control is maintained by continuously monitoring the atmospheric-beryllium levels. This is done in a variety of ways, consisting essentially of sucking a known volume of air through a filter paper and analysing the entrapped material by means of a refined spectrographic technique capable of detecting as little as 5×10^{-8} g of beryllium. With such rigorously controlled working conditions, atmospheric contamination can be kept below specified tolerance levels and the metal processed without risk.

BERYLLIUM RESEARCH

The main aims of beryllium research are to give the metal more uniform room-temperature ductility and to improve its workability at higher temperatures by eliminating minima in the ductility/temperature curve.

In order to make beryllium satisfactorily ductile at room temperature the principal difficulty to be overcome is that of cleavage fracture on the basal plane. Reasons for the observed behaviour may be of two basic kinds, one deriving from the fundamental properties of metallic beryllium itself and one due to impurities. If the fault lies with the metal, it must be assumed that the interatomic bonding within the planes on which premature fracture occurs is very much stronger than that between them, that is, the crystal has a layered structure similar to that of graphite, involving partially directed bonds and implying some departure from ideal metallic behaviour. If this type of explanation is true, the only way in which the situation can be improved is to alloy the beryllium with a metal which can enter into solid solution with it and by thus changing its electronic structure modify the type of bonding existing in the metal crystal, or even the crystal structure itself. If, on the other hand, impurities are playing a major role—as they could, for instance, by raising the temperature of the ductile/brittle transition undergone by many metals at low temperatures or by forming oxide layers at the grain boundaries—the choice lies between removing the impurities altogether or eliminating them from the metal crystal by a precipitation process, such as can be achieved by taking advantage of a change of solubility with temperature, or alternatively by adding an alloying element which will combine preferentially with the impurity.

Although room-temperature brittleness in beryllium could be an inherent property of the metal, it is difficult to envisage any fundamental characteristic of the beryllium crystal which could give rise to the complex ductility/temperature relationship observed, and there is growing confidence that this is an impurity effect. Again the choice of remedy lies between purification and precipitation of the impurities by heat-treatment or alloying.

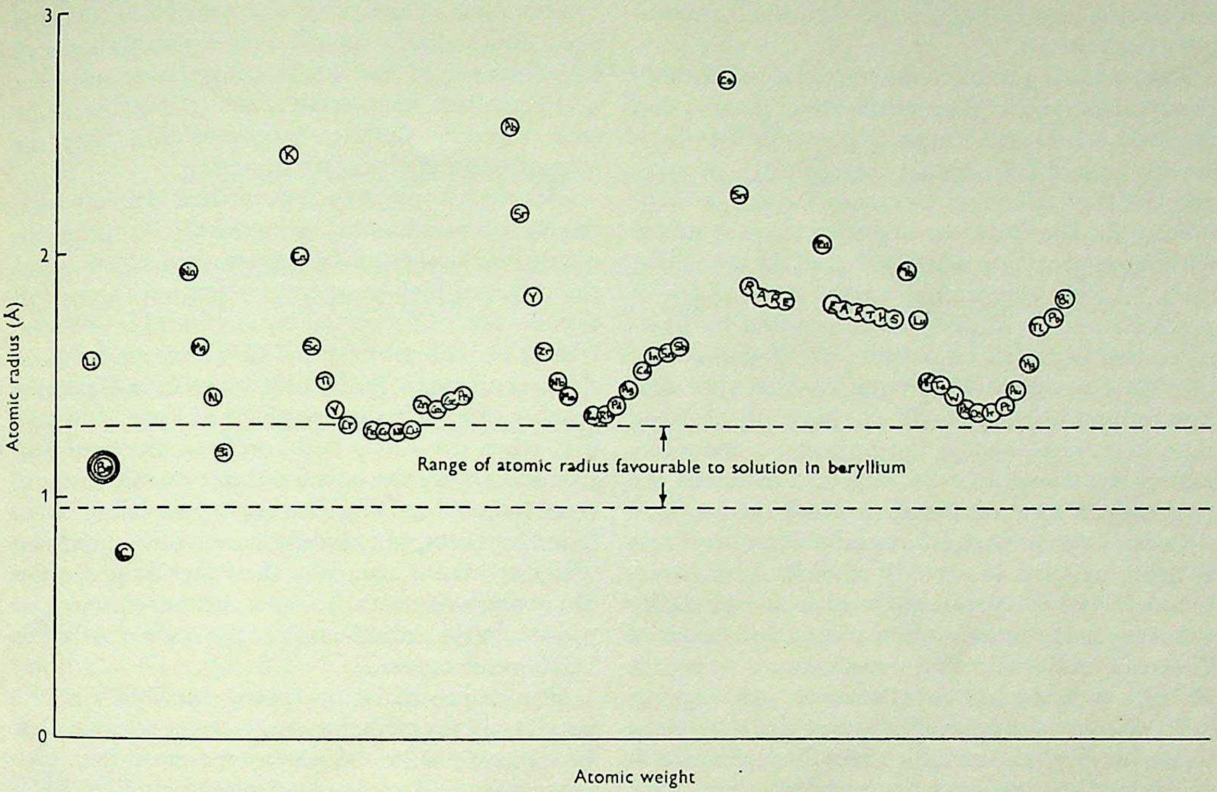


FIGURE 4 – *Atomic radii of the metallic elements. Only silicon, iron, cobalt, nickel, and copper fall within the range favourable to solution in beryllium.*

Consideration of the atomic properties of beryllium suggests that attempts to modify the electronic nature of the metal are unlikely to be rewarding. The beryllium atom is very small, having an atomic radius of 1·125 Å. Normally it is observed that extensive solid-solubility between one metal and another occurs only when the atomic radius of the solute element differs from that of the solvent element by not more than 15 per cent of the atomic radius of the latter. Consequently little solubility in beryllium can be expected from the majority of elements, since all but a few have atomic radii well outside the favourable range (figure 4). The tendency will rather be to form intermetallic compounds, more particularly as beryllium is one of the more strongly electropositive metals—a fact which favours compound formation with electronegative elements. The chance of dissolving a proportion of a substitutional alloying element in beryllium sufficient to modify appreciably the electronic characteristics of the metal seems thus to be very small. In fact only copper, palladium, nickel, cobalt, and silver have any appreciable solubility in beryllium, and even for these elements the range of solid solution is too limited to offer much hope of influencing the fundamental nature of the metal. Attempts have been made to stabilize, at lower temperatures, the body-centred cubic structure that beryllium is reported to adopt at temperatures near its melting point, by adding the more promising alloying elements, and though it might have been expected that without more extensive solid-solubility than has yet been found there would be little likelihood of success, recent work indicates that it may in fact be possible to retain the high-temperature structure at temperatures approaching the range suitable for hot working. Some consideration is also being given to the possibility of stabilizing this structure by the application of high pressures.

Any aspects of the behaviour of beryllium which are due to impurities have a considerably greater chance of being modified. Methods of purification under investigation include distillation, iodide refining, zone refining, electrolysis, and electron-beam melting. These have not yet produced a form of beryllium which is free of processing difficulties, and it is probable that the embrittling mechanisms operating involve very small quantities of impurities.

In attempting to solve the impurity problem by precipitation rather than purification, the fact that the solid solubility of most elements in beryllium is very limited helps rather than hinders, as small additions of alloying elements having a high affinity for the common impurities can be added without appreciable modification of the beryllium itself. A full exploration of the possibilities of precipitation by alloying additions and by heat-treatment forms an important part of present-day beryllium research. Hydrogen, oxygen, nitrogen, iron, aluminium, and silicon have all at some time been suggested as the impurity responsible, and current work may be expected to define the part played by these elements more accurately.

Some of the research effort available for beryllium is being devoted to a study of dislocation movements in beryllium crystals, which should clarify the reasons of the room-temperature brittleness of the metal and establish the mechanisms by means of which impurity elements can influence its deformation behaviour. The knowledge that, given a satisfactory grain-structure, beryllium can exhibit good ductility has also encouraged efforts to bring about the desired structure in material as cast.

THE FUTURE OF BERYLLIUM

Though it has required the stimulus of atomic energy to produce the effort necessary for the full industrial development of beryllium, the metal might well find other applications. The most obvious possibility is in the aircraft industry, which could take advantage of its extreme lightness and rigidity. The same qualities may also find some applications in missiles. Less dramatic uses are found in gyroscopic components, which require very high rigidity plus dimensional stability, and also in X-ray sets, for which beryllium, being transparent to X-rays, is an ideal window-material. This last application was the first industrial use of metallic beryllium.

As with so many other of the less common metals, the future of beryllium depends on its availability, its cost, and the extent to which its fundamental disadvantages can be overcome. With its unique combination of desirable characteristics, there is little doubt that a ductile and workable form of beryllium would find important applications, even if it remained very costly; but it remains to be seen whether ductile beryllium will ever be offered as an engineering material.

Although it might be argued that, if a very high degree of purification is necessary in order to obtain satisfactory room-temperature ductility and the required high-temperature properties, the full exploitation of beryllium by engineers is unlikely, it must be remembered that in the case of hafnium the dependence of its ductility on such an expensive refining process has not prevented its use. It is true that when the costly iodide-refined titanium and zirconium were the only available ductile forms of these metals no engineering applications were found for them, but the engineering developments which created the need for their specific properties did not take place until after a cheaper commercial process for the manufacture of the ductile materials had been discovered.

The nature of its processing inevitably makes beryllium expensive, but increased production can be expected substantially to reduce price. Ten years ago the American price for beryllium billet was over $100 per lb, but a demand for about 50 tons a year quickly reduced this to $47 per lb. Comment on the trends of British prices would as yet be premature, but they are not expected to exceed current American prices. In making comparisons of cost it must be remembered that beryllium is appreciably lighter than other metals, and that the figure used for most comparative purposes should be the cost per unit volume; on this basis beryllium appears in a much more favourable light. Moreover, full-scale industrial exploitation would bring about even greater price reductions than have so far occurred.

What is now needed is a better understanding of the behaviour of the metal itself and of its toxicity so that processing can be simplified and cheapened. With advances in metallurgical and medical knowledge it can be hoped that within the next few years it will become possible to assess the future of beryllium far more accurately than can be done today.

REFERENCES

[1] KAUFMANN, A. R., GORDON, P., and LILLIE, D. W. *Trans. Amer. Soc. Metals*, 42, 786, 1950.

[2] BROWN, A. B., MORROW, F., and MARTIN, A. J. *Nature, Lond.*, 187, 494, 1960.

The transport of organic substances in plants

A. L. KURSANOV

Although it has long been established that the phloem and the xylem are the media by which organic substances are transported in plants, the nature of the mechanisms involved is still obscure. In this article the author gives a general review of present knowledge and its theoretical interpretation, with special reference to recent research in the Soviet Union. He concludes with an interesting discussion of the exchange of products between mature leaves.

The old theory that substances are transported inside plants in the cells of the xylem and phloem remains unquestioned today, but many debatable problems arise when the nature of the movements, and especially that of the descending flow, is considered. During the past decade many valuable facts have been established by tracer techniques, but two conflicting theories still remain in the field. According to the first, proposed by E. Münch in 1932 [1], a solution flowing in the sieve cells carries the assimilates passively downwards, the motive force arising from the difference of turgor pressure between the two ends of the system. According to the second theory, there is no mass flow of solution, but an activated movement of molecules. The energy required is supposed to result from the metabolism of the sieve cells or their companion cells.

Neither theory has, as yet, decisive evidence in its favour. The best support for the first theory comes, in my opinion, from the work of T. E. Mittler [2] and of K. Mothes and L. Engelbrecht [3]. They have shown that after a sieve cell is punctured by an aphid it continues to exude a sucrose solution for a long period, until the exudate exceeds by many thousands of times the volume and sucrose content of a single sieve cell. This may be a pathological disturbance of the normal system, but it does suggest that a free flow of sugar solution is possible through the sequence of sieve cells.

The mass-flow hypothesis encounters a number of difficulties. Cytological observations do not normally reveal connecting channels (plasmodesmata) along which a continuous flow of solution could take place from one sieve cell to another. Moreover, if such channels exist, they must be so narrow that the resistance to flow would be far greater than can be reconciled with the rates indicated by isotopic studies. Another difficulty is that materials can move simultaneously at different rates [4, 5, 6] and even in opposite directions [7].

For these reasons we may turn to the activated-transport hypothesis, but it is still not clear how the energy of metabolism is utilized. More detailed study of the metabolism of the phloem tissue is needed, and my colleagues and I have been working in this field for a number of years. Studies carried out by M. N. Zaprometov in 1949 [8, 9] showed that amino acids moved from a solution through the cut ends of wheat stems into the ears in 50 to 100 times greater quantities than could be accounted for by carriage in the transpiration stream. This observation suggested that the transport of organic materials might be dependent on metabolism.

M. V. Turkina [10], using ^{14}C, showed in 1954 that the principal substance transported in the conducting strands of sugar beet is sucrose; a similar result had already been recorded by L. P. Vernon and S. Aronoff [11] for soya (*Glycine soja*). The results further indicated that the root of the sugar beet cannot itself synthesize sucrose [12, 13], but receives it ready-made from the leaves.

The predominance of sucrose as a transport substance has since been shown for many other species. It is especially evident in experiments carried out by M. K. Chailakhyan and myself [14] with reciprocal grafts of sunflower (*Helianthus annuus*) and Jerusalem artichoke (*Helianthus tuberosus*) arranged as in figure 1. The artichoke forms large quantities of fructosans, from the simplest to high-polymer types, while the sunflower is typified by the formation of starch and monosaccharides. When a leaf of the top component of the graft was fed with $^{14}CO_2$, it became evident that the carbohydrates synthesized travelled to the roots as sucrose irrespective of whether they were first formed in leaves of artichoke or sunflower, and irrespective of which component of the mixed plant was being traversed. The high proportion of the ^{14}C found in sucrose when the last of the different sections was analysed is also indicated in figure 1. The role of sucrose as a transport material is also well illustrated in the experiments of J. Edelman, S. I. Shibko, and A. J. Keys [15]. When ^{14}C-labelled glucose is absorbed by the scutellum of wheat or barley, it is first rapidly converted to sucrose and only then is it conveyed to the shoots and roots. In a few species sucrose appears to be replaced by analogous galactosides and in part by raffinose [16, 17]. This does not contradict the central role of sucrose and even perhaps extends it, for it shows that galactose residues can be transported when attached to sucrose.

Amino acids, organic acids, and probably many other substances move in the phloem in addition to sugars. A detailed study has been made of the organic and amino acids in the descending current. For example, M. I. Brovchenko [5] has shown in our laboratory that, when a rhubarb (*Rheum* sp.) leaf is partly covered with a chamber containing $^{14}CO_2$, not only labelled sugars but labelled organic and amino acids can be detected in the adjacent veins within 3–4 minutes (figure 2). The amino acids generally proved to be more mobile than the organic acids, but the most significant point is that the assimilates entered the conducting tissues in proportions different from those in the photosynthesizing cells. In rhubarb the most mobile acids were malate and citrate, and among the amino acids threonine, serine, and alanine. Others, for example succinic and fumaric acids, either did not enter the conducting strands at all or did so only very slowly (figure 3). The composition of the descending solution may vary with the stage of growth of the plant. In the early autumn, the rate of movement of proline and aspartic acid diminishes, whereas that of threonine and serine remains the same and that of γ-aminobutyric acid even increases. Soya plants showed

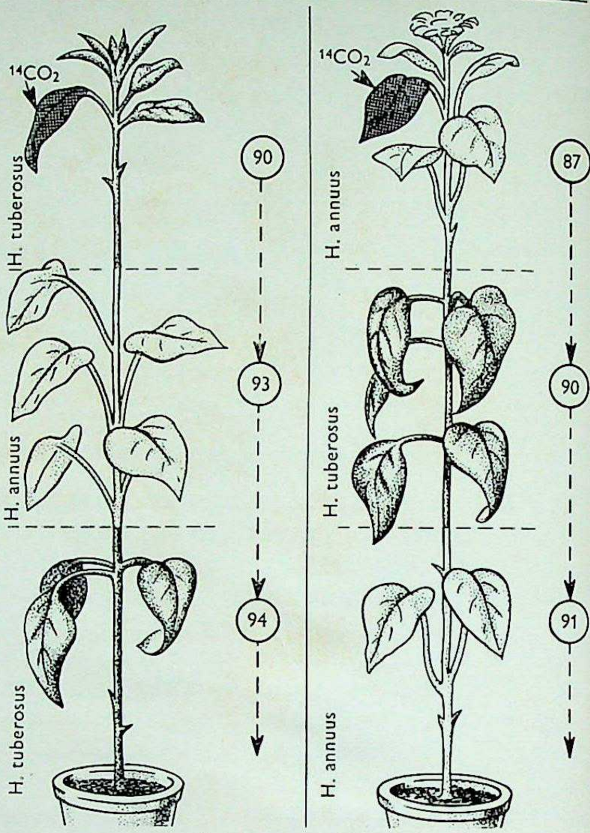


FIGURE 1 – *Grafts of sunflower and artichoke fed with $^{14}CO_2$ through the heavily shaded leaf only. The figures in circles show the percentages of the total ^{14}C in each zone recovered as sucrose.*

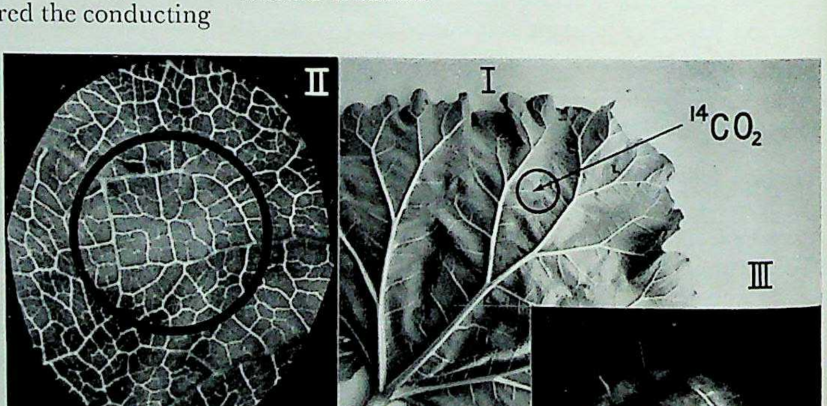


FIGURE 2 – *Underside of rhubarb leaf. In I, the dark circle shows the area to which $^{14}CO_2$ was fed; II is a portion of the same enlarged, showing the smaller veins. III is a radio-autograph showing, after a few minutes, the distribution of ^{14}C in the mesophyll* (a), *veins* (d), *and veins outside the area of supply* (b).

63440 R 505 L64E 07.7826

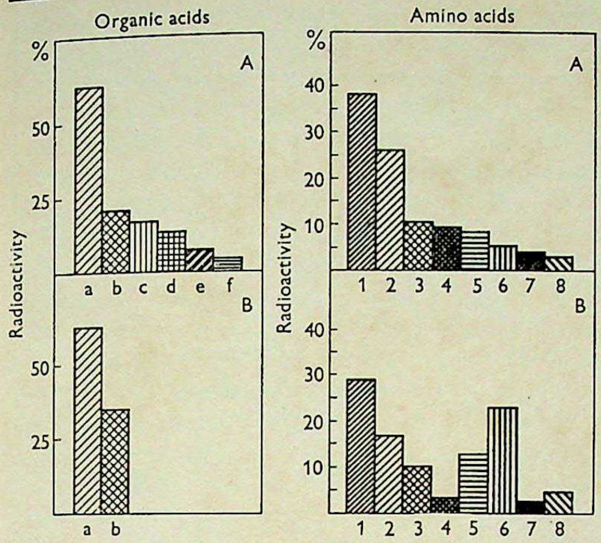


FIGURE 3 – *Distribution of ^{14}C in rhubarb leaf after feeding $^{14}CO_2$ to a restricted area. A, ^{14}C recovered from mesophyll; B, ^{14}C recovered from veins (conducting strands). The following organic acids were examined:* (a) *malic,* (b) *citric,* (c) *succinic,* (d) *fumaric,* (e) *tartaric, and* (f) *oxalic. The amino acids* (1) *serine,* (2) *alanine,* (3) *glycine,* (4) *aspartic,* (5) *glutamic,* (6) *threonine,* (7) *γ-aminobutyric, and* (8) *proline were also tested.*

similar results (C. D. Nelson and P. R. Gorham [6]). A drop of solution containing radioactive amino acids was placed on a leaf stalk from which the blade had been cut away. The different amino acids were conveyed within the plant at different rates and even in different directions. In 17-day-old plants, serine, which accumulates rapidly in young tissues, was particularly mobile, whereas asparagine and glutamine hardly moved. In older plants these relations were reversed.

The rate of movement of ^{14}C-labelled photosynthetic products has proved to be unexpectedly fast. According to the findings of M. V. Turkina and I. Dubinina [18] and N. A. Pristupa [19], it varies in grasses from 40 to 120 cm/hr and similar values have been obtained by other investigators. These rates are much in excess of any value that could be expected from simple diffusion.

Nelson and Gorham have observed still faster rates for amino acids, which are conveyed through 20- to 24-day-old soya plants at 370–1370 cm/hr and sometimes even at 3000 cm/hr. Their technique is, however, subject to the danger that some of the droplet applied to the cut end of the leaf stalk may be sucked into the opened xylem vessels. Later experiments by C. D. Nelson, H. J. Perkins, and P. R. Gorham [20], in which soya leaves were fed with $^{14}CO_2$ showed that a small part of the assimilates containing ^{14}C moved towards the roots at over 5000 cm/hr and that this was followed by the main mass of assimilates at about 100 cm/hr. It may be concluded that there are two separate mechanisms for the transport of organic materials in the phloem. Even the observation of such speeds as 100 cm/hr creates difficulties for the theory of mass flow, owing to the extreme fineness of the channels connecting phloem cells and the large differences of osmotic pressure that would therefore be required. Even for the metabolic theory they are a serious test, since they would require considerable expenditure of energy in the conducting tissue.

Turkina [21] has made an important contribution from this point of view by showing that the conducting tissues of sugar beet, *Plantago* sp., and *Caragana arborescens* have high respiratory intensities. These and similar results are summarized in Table 1. H. Ziegler [24] was able to divide the bundles of *Heracleum* into xylem and phloem; he showed that respiration was fast in both tissues.

It has been shown by Turkina and Dubinina [25] that the respiratory quotient (CO_2/O_2) of

TABLE I

Respiration rates in fibrovascular bundles and other tissues

Plants	Respiration (μl O_2 per gram of fresh tissue)			Authors
	Fibro-vascular bundles	Leaves	Petioles (without bundles)	
Sugar beet	572	416	100	Kursanov and Turkina, 1952
Sugar beet	462	232	95	Tsao-Tsung-Hsun and Liu Chin Yu, [22]
Plantago major	820	374·8	228	Kursanov and Turkina, 1952
Plantago major	500	—	146	Willenbrink, [23]
Primula beesiana	309	—	52	Willenbrink, [23]
Heracleum mantegazzianum	230	—	32	Ziegler, 1958

fibrovascular bundles is close to unity, that is, that their respiration proceeds mainly at the expense of carbohydrates. In the conducting tissues, cytochrome oxidase predominates over other oxidases, an indication of the nature of the terminal stages of their respiration.

The analyses of the phloem juice of *Robinia pseudacacia* carried out by H. Wanner [26] are frequently cited as evidence for the inactivity of sieve-tube cells, which are known to be without a nucleus after a certain age. It would, however, hardly be correct to identify the enzymatic activities of a phloem exudate from damaged bark with the enzyme complement of the protoplasm itself. Furthermore, the sieve tubes are in close connection with their companion cells and appear to form a single metabolic system.

We now have detailed knowledge of the enzymes of the vascular bundles in sugar beet (Table II). This shows that the bundles are able to activate hexoses via the hexokinase reaction. The presence of active aldolase and succinic dehydrogenase indicates that respiration is likely to proceed by way of glycolysis and the Krebs cycle. Invertase, on the other hand, is almost absent and this accounts for the relative stability of sucrose in the conducting strands.

Nevertheless, some of the sugar moving in the phloem is transformed. O. A. Pavlinova [27] showed that ^{14}C-labelled fructose moving along the conducting strands of sugar beet is converted rather rapidly into sucrose labelled equally in the fructose and glucose moieties. No labelled glucose was detectable, indicating a rapid isomerization of hexoses and utilization of their phosphoric esters for sucrose synthesis. This conclusion was confirmed by the presence of uridine monophosphate and uridine diphosphate in fair quantities in the vascular bundles. Turkina [28], using isolated bundles, showed that labelled sucrose utilized in respiration is converted within the conducting tissues to a mixture of acids including pyruvic, hydroxypyruvic, α-ketoglutaric, oxalacetic, and glyoxylic acids. These keto acids lead, by amination and transamination, to a series of amino acids that were also identified.

TABLE II

Enzymes of glycolysis and oxidation in conducting tissues (sugar beet)

Enzymes	Activity	Authors
Aldolase	+++	Pavlinova
Apyrase	++	Pavlinova
Cytochrome oxydase	++++	Turkina and Dubinina
Phosphatase (acid)	++	Turkina and Dubinina
Phosphorylase	+	Turkina and Dubinina
α-Galactosidase	+	Pavlinova
Hexokinase	+++	Pavlinova
Phosphohexose isomerase	+	Pavlinova
Invertase	+	Turkina and Dubinina
Succinic dehydrogenase	++	Ziegler

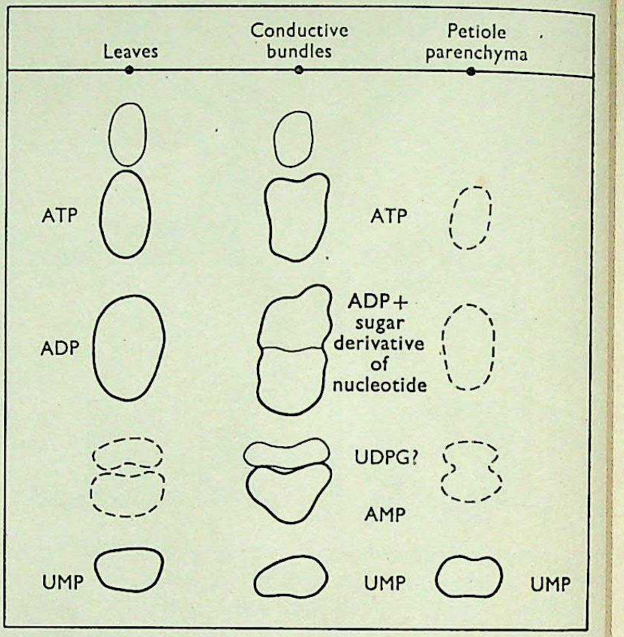


FIGURE 4 – *The acid-soluble nucleotides in the conducting tissues of sugar beet: AMP adenosine monophosphate, ADP adenosine diphosphate, ATP adenosine triphosphate, UMP uridine monophosphate, UDPG uridine diphosphate glucose.*

The principal transformations of sucrose thus occur by way of glycolysis and the Krebs cycle, giving rise to many intermediate and secondary products. The consumption of sucrose in these reactions is probably slight compared with the total amount passing through the conducting strands in the same time. Nevertheless, the intense respiration of the conducting tissues permits one to conclude that enough labile phosphate bonds (ATP) are formed for the purpose. Indirect support comes from the high activity of hexokinase and the relative abundance of nucleotides and other organo-phosphorus compounds.

Figure 4 shows chromatograms of nucleotides obtained from the fibrovascular bundles of sugar beet. Analogous chromatograms are given for the leaf blades and parenchyma of the petioles of the same plants. Comparison shows that the bundles are far richer in nucleotides than the parenchyma surrounding them or than the laminae.

We can now give a fairly detailed picture of the metabolism of the conducting tissues, but the question remains as to whether the energy liberated sets the machinery of transport in motion. The answer should apparently be in the affirmative, though we do not yet know just how the energy is utilized. It has been shown that the transport of organic materials is much retarded, or even prevented altogether, when the fibrovascular bundles are poisoned by respiratory inhibitors, such as carbon monoxide, cyanide, dinitrophenol, arsenite, azide, or iodoacetate. A similar result may also come from local cooling of the conducting tissue to 2–5° C, as was shown by O. F. Curtis in 1919. All this suggests that active metabolism in the conducting tissues is a prerequisite for normal transport of organic substances through them. The same conclusion has been reached by Nelson and Gorham [6, 29] who have shown that the transport of sucrose and certain amino acids is inhibited by cyanide, or by high temperatures: whereas the movement of sucrose and arginine, for instance, ceased entirely in damaged phloem, that of monosaocharides and asparagine continued. The latter movement may be explained by solution entering the vessel cavities of the cut xylem, a possibility not excluded by the method used.

The principal transport compounds, especially sucrose, are not sucked into the vessels because the protoplasm of the surrounding cells has a higher 'affinity' for them and so retains them. New experiments supporting this view have been carried out by Turkina [30], who has shown that fibrovascular bundles isolated from leaves of sugar beet and other plants have a capacity for accumulating sucrose greatly in excess of that of other tissues. Glucose and glycine are absorbed by the conducting strands much less, and selective absorption of sucrose may be achieved by placing conducting tissues in mixed solutions (Table III). This property is possessed by both the phloem and the xylem of the conducting strands. The accumulation of sucrose is strongly inhibited by cyanide and dinitrophenol (Table IV). We still know very little about the affinity of the conducting tissues for other materials, such as some of the amino acids; but it already appears plausible that the difference of affinity is one of the reasons for the selective uptake of sucrose and certain amino acids out of the photosynthesizing cells into the channel of flow. It may also explain why fibrovascular bundles remain rich in sucrose throughout their length while the surrounding tissues contain mainly monosaccharides.

TABLE III

Comparative absorption of ^{14}C-sucrose, ^{14}C-glucose, and ^{14}C-glycine-1 by the conducting bundles of sugar beet, September 1959. (*In counts per minute per gram fresh weight*)

Substance	Specific activity of 1 ml solution	Absorption (1 hour)		Ratio *a*/*b*
		(*a*) Conducting bundles	(*b*) Petiole parenchyma	
Sucrose . .	103 330	20 610	2880	7·2
Glucose . .	109 800	7450	3980	1·8
Glycine . .	108 870	4820	2940	1·6

TABLE IV

Influence of respiratory poisons on the absorption of sucrose by the conducting bundles of sugar beet

Duration of experiment	Solution	Absorption Counts per minute per grain	Percentage inhibition
1 hour	Sucrose	25 222	64·0
	Sucrose + KCN (10^{-3} M)	9073	
2 hours	Sucrose	52 937	51·2
	Sucrose + DNP (10^{-4} M)	25 860	

The transport of organic materials over long distances is dependent not only on the metabolism of the conducting strands but on the activities of the organs at either end. Assimilates in mesophyll cells are delivered into the conducting system very rapidly and usually independently of the concentration gradient [29, 31]. Any idea that this speed is due simply to diffusion must be given up, and an explanation sought in terms of glandular secretion or of a selective accumulation at the expense of the metabolic energy of the conducting tissues themselves. The work of Turkina just mentioned makes an active role of the conducting tissues themselves seem highly probable, but does not exclude the other possibility.

The mechanism of entry of the products of photosynthesis into the conducting system has not

TABLE V

Influence of ATP on the flow of assimilates (^{14}C) *from the mesophyll into the conducting bundles in sugar beet* (*Counts per minute per gram*)

	June		September	
	H_2O	ATP	H_2O	ATP
Mesophyll	224 490	200 040	94 760	112 540
Conducting bundles (cm)				
3–4	2470	3110	300	3450
5–6	1300	2170	230	1160
7–8	720	1130	110	740
9–10	350	870	80	280
11–12	340	530	—	—
13–14	180	320	—	—
Total entering the bundles	5360	8130	720	5630

been sufficiently studied, but it would seem that active metabolism is an essential condition. From analogy with the absorption of nutrients through the walls of the intestine, the passage of sugars from photosynthesizing cells into the conducting tissues is often thought to depend on phosphorylation. As yet there is little direct proof of this. I should like to call attention to several experiments recently carried out in our laboratory by Brovchenko. He found that enriching sugar beet leaves with adenosine triphosphate (ATP) caused a perceptible acceleration in the release of assimilates formed after feeding $^{14}CO_2$ for five minutes into the conducting strands, and a faster movement towards the root (Table V). The observations were made in the dark during a period of 25 minutes, that is, under conditions which precluded the photosynthetic formation of ATP. We are at present studying the composition of the assimilates entering the conducting system in the presence of the added ATP.

The movements of organic substances through the plant are apparently very complicated; both composition and direction of flow vary considerably for a variety of reasons. In 1953, N. A. Pristupa obtained radio-autographs from 22-day-old pumpkin plants which showed that the bulk of ^{14}C-labelled leaf assimilates moved to the root. In later experiments [19], 50 per cent was found to have reached the roots and part returned to the above-ground organs in the transpiration stream after changing their form. Thus arose the idea of a cycle of organic matter in plants.

The preferential movement of assimilates to the roots has been corroborated by many other investigators, working with annual plants up to the time of their flowering; as a plant ages the distribution of assimilates changes. This has been vividly demonstrated by I. F. Belikov [32] with soya plants in the neighbourhood of Vladivostok. Using $^{14}CO_2$ he found that the leaves of mature plants supply specific organs, or zones, with their assimilates. Low-level leaves send their products mainly to the roots but, as the fruit develops, ever increasing numbers of leaves pass their assimilates to it, and at length even those at the lower levels follow suit, and cease almost entirely to feed the roots. At this stage the plant does not react to additional

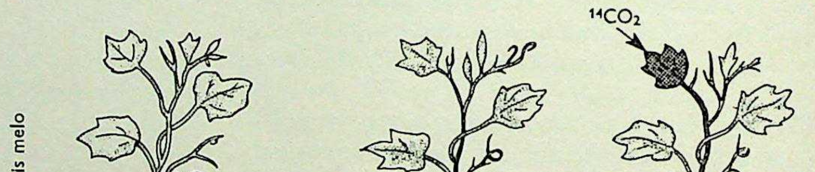


FIGURE 5 – *Graft of* Cucumis melo *on* Cucurbita ficifolia.

fertilizer, since its roots cannot form enough ion-acceptors.

Young leaves that have not completed their growth are able to photosynthesize, but do not yield their assimilates to other parts of the plant. On the contrary, they continue for a long time to receive products of photosynthesis from the nearest mature leaves. Exchange of products between mature leaves at different levels virtually ceases and it is not possible to bring about a flow of assimilates to a mature leaf even by prolonged darkening. This severe rule—the absence of mutual aid between mature leaves—may be of considerable biological importance, since it relieves the plant of the necessity of supporting leaves that are in a position unfavourable for photosynthesis.

Recent studies with tobacco plants [33] also show that ^{14}C-labelled assimilates from mature leaves continue to enter young ones for a considerable period. At the end of its growing period, a leaf absorbs and 'exports' assimilates simultaneously; in its mature state, export becomes predominant. At this stage, the flow of assimilates to the young leaves possibly goes up the xylem, into which a part of the assimilates penetrate from the phloem. This may explain why labelled assimilates could be found only in leaves above the one fed with $^{14}CO_2$ and were not found at all in the lower leaves, even though the bulk of the photosynthetic products moved down to the roots.

The distribution of assimilates in plants is probably far more complex than is suggested by these experiments. H. C. M. De Stigter [34] has shown that it is possible to graft musk-melon (*Cucumis melo*) on to *Cucurbita ficifolia* only if a few leaves are left on the stock, since the leaves of the scion do not feed the alien roots. It is enough to graft a single leaf of the stock species on to the top of the scion to start all the leaves of the scion sending assimilates down to the stock (figure 5). The nature of this effect is obscure.

It seems that the movement of organic substances in plants is multifariously regulated by means of the metabolism of all parts of the plant, and I think that the conducting cells of the phloem are no exception. We believe that more attention should be paid to the problem of the relation of metabolism to the transport of organic materials, and it is this aspect of the matter that we are trying to develop in our laboratory.

REFERENCES

[1] Münch, E. 'Die Stoffbewegungen in der Pflanze.' Jena. 1932.
[2] Mittler, T. E. *Nature, Lond.*, **172**, 207, 1953.
[3] Mothes, K. and Engelbrecht, L. *Proc. All-Union Bot. Congr.*, Leningrad. 1957.
[4] Biddulf, O. and Cory, R. *Plant Physiol.*, **32**, 609, 1957.
[5] Kursanov, A. L., Brovchenko, M. I. and Pariiskaya, A. Физиол. Растении, **6**, 527, 1959.
[6] Nelson, C. D. and Gorham, P. R. *Canad. J. Bot.*, **37**, 431, 1959.
[7] Chen, S. L. *Amer. J. Bot.*, **38**, 203, 1951.
[8] Kursanov, A. L. and Zaprometov, M. N. Докл. Акад. Наук, С.С.С.Р., **68**, 1113, 1949.
[9] *Idem. Ibid.*, **69**, 89, 1949.
[10] Kursanov, A, L. and Turkina, M. V. *Ibid.*, **95**, 885, 1954.
[11] Vernon, L. P. and Aronoff, S. *Arch. Biochem. Biophys.*, **36**, 383, 1952.
[12] Kursanov, A. L. and Pavlinova, O. A. Биохимия, **17**, 446, 1952.
[13] Pavlinova, O. A. *Ibid.*, **19**, 364, 1954.
[14] Kursanov, A. L., Chailakhyan, M. K., Pavlinova, O. A., Turkina, M. V., and Brovchenko, M. I. Физиол. Растении, **5**, 3, 1958.
[15] Edelman, J., Shibko, S. I. and Keys, A. J. *J. exp. Bot.*, **10**, 178, 1959.
[16] Zimmerman, M. H. *Plant Physiol.*, **33**, 213, 1958.
[17] Pristupa, P. A. Физиол. Растении, **6**, 30, 1949.
[18] Kursanov, A. L., Turkina, M. V., and Dubinina, I. Докл. Акад. Наук, С.С.С.Р., **93**, 1115, 1953.
[19] Kursanov, A. L. and Pristupa, N. A. Физиол. Растении, **4**, 417, 1957.
[20] Nelson, C. D., Perkins, H. J., and Gorham, P. R. *Canad. J. Bot.*, **37**, 1181, 1959.
[21] Kursanov, A. L. and Turkina, M. V. Докл. Акад. Наук, С.С.С.Р., **84**, 1073, 1952.
[22] Tsao Tsung Hsun and Liu Chin Yu. *Acta Bot. Sin.*, **6**, 2699, 1957.
[23] Willenbrink, J. *Planta*, **48**, 3, 1957.
[24] Ziegler, H. *Ibid.*, **51**, 186, 1958.
[25] Turkina, M. V. and Dubinina, I. Докл. Акад. Наук, С.С.С.Р., **95**, 199, 1954.
[26] Wanner, H. *Ber. schweiz. bot. Ges.*, **63**, 201, 1953.
[27] Pavlinova, O. A. Физиол. Растении, **2**, 378, 1955.
[28] Turkina, M. V. *Ibid.*, **6**, 709, 1959.
[29] Nelson, C. D. and Gorham, P. R. *Canad. J. Bot.*, **35**, 703, 1957.
[30] Turkina, M. V. Биохимия, **19**, 357, 1954.
[31] Ziegler, H. *Planta*, **47**, 447, 1956.
[32] Belikov, I. F. Физиол. Растении, **2**, 354, 1955.
[33] Jones, H., Martin, R., and Pork, H. K. *Ann. Bot. Lond.*, N.S., **23**, 493, 1959.
[34] De Stigter, H. C. M. *Proc. IX Inter. Bot. Congr.*, **2**, 90, 1959.

Polarography

G. C. BARKER and G. W. C. MILNER

Diffusion near any stationary electrode is an ill-defined process and it is impossible to obtain reliable information about the composition of the solution from a study of the variations in current resulting from alterations in the potential of the electrode. This and other difficulties are avoided by using a dropping-mercury electrode, and polarography, the analytical method made possible by its use, is extremely rapid, and is applicable to very low concentrations of reducible or oxidizable substances.

When a potential difference exists between two stationary metal electrodes immersed in a solution containing oxidizable or reducible substances, reactions normally occur at the electrodes in which the cathode gives electrons to ions or molecules in the solution and the anode accepts electrons: a current flows through the cell. The size of this current depends on the composition of the solution and on the applied voltage, and some information about the composition of the solution can be obtained from a study of the relation between voltage and current. This information is only qualitative in the case of a simple cell. The current/voltage curves are not reproducible, because of the effects of the accumulation of the products of electrolysis on or near the electrodes, and because the cell current may be affected by convection currents within the solution. These difficulties are evaded in polarography—a technique originally described by J. Heyrovsky in 1922—by the use of a dropping-mercury electrode and of a 'supporting' electrolyte (see later): the current then varies in a stepwise fasion as the potential difference between the electrodes is increased; a solution containing several oxidizable and reducible substances will give a current/voltage curve from which it is possible to identify the various species and to determine their concentrations.

TECHNIQUE

The dropping-mercury electrode usually consists of a length of glass capillary tube ($\sim 0{\cdot}05$ mm diameter) that is attached to a reservoir of mercury by means of flexible tubing. The drops of mercury issue regularly from the tip, and fall at the rate of one every 3–6 seconds. This electrode is normally made the cathode of the polarographic cell; the anode is usually a large pool of mercury (figure 1). The dropping-mercury electrode has the advantage that contamination of the electrode surface by reaction products is avoided, as the surface is continuously renewed. It also enables polarography to be used with acid solutions, as the reduction of hydrogen ions takes place only with difficulty at a mercury surface.

To illustrate the principle of the method, suppose the electrodes to be immersed in a solution containing a few milligrams of cadmium chloride per litre. The cadmium ions near to the cathode are reduced to the metal when the cathode's potential is about 1·0 V lower than that of the anode. The cadmium metal dissolves in the mercury, forming an amalgam that is removed when the mercury drop falls. The cadmium ions, on reduction at the cathode, give up their charge; current flows between the two electrodes; and an equivalent amount of mercurous chloride is formed at the mercury-pool anode. The removal of cadmium ions from the solution near the mercury drop results in the formation of a radial concentration-gradient in that region, and cadmium ions thus tend to diffuse towards the cathode. The radial flux of cadmium ions at the surface of the mercury drop at any time is dependent on the concentration gradient, but in the case of a dilute solution it may also be affected by electro-migration caused by potential gradients within the solution. Such electro-migration is undesirable in polarography, and it is suppressed by adding to the solution a relatively large amount of an indifferent salt, known as the supporting electrolyte. The current then observed at the cathode is due only to the process of diffusion.

When the fixed voltage is replaced by a slowly increasing negative voltage, no current flows through the cell at values of the voltage below that for the reduction of the cadmium ions. After attainment of this voltage, however, reduction begins, and the concentration of cadmium ions in the solution at the electrode surface rapidly falls to a negligibly small value. As the current is simply related to the rate of arrival of cadmium

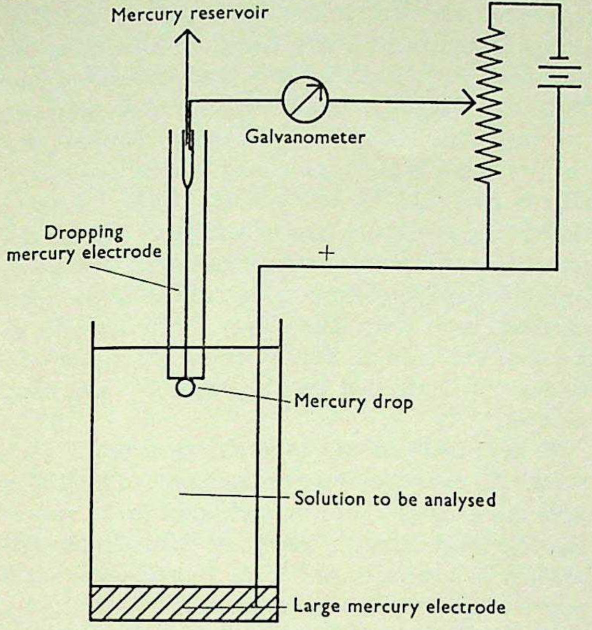


FIGURE 1 – *Polarographic cell (schematic).*

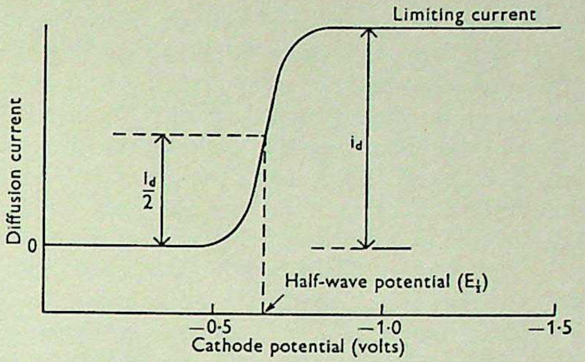


FIGURE 2 – *Idealized polarogram for a solution of a cadmium salt.*

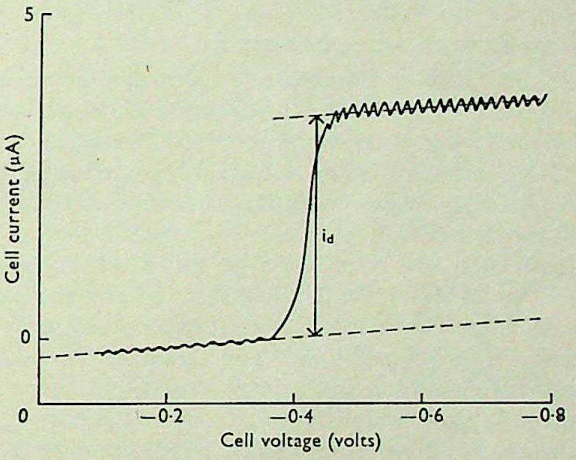


FIGURE 3 – *Typical polarogram for lead ion in potassium chloride solution.*

ions at the electrode surface, and this in turn is proportional to the concentration gradient, the current rises rapidly to a value that is not noticeably affected by further increases in the applied voltage (figure 2). This limiting value of the current (i_d) is generally referred to as the limiting diffusion-current and, provided the reduction process is diffusion-controlled, this current is proportional to the concentration of the reducible species in the bulk of the solution; this fact is the basis of the application of polarography to quantitative analysis. The limiting current varies somewhat during the life of a single mercury drop, because the drop is growing, and normally one measures its average value. This is influenced by such factors as the diffusion coefficient of the reducible species, the rate of flow of mercury, and the 'drop time'; the dependence of the current on all these factors is well understood [1, 2]. These factors are usually held constant in practical polarography, and concentrations are determined by adding known amounts of the species in question to the solution and studying the effect of these additions on the limiting diffusion-current.

A current/voltage curve of an electrode reaction is called a polarographic wave (figure 2), and the potential of the dropping-mercury electrode (with respect to a suitable reference-electrode) corresponding to the point on the wave at which the diffusion current is one-half of the limiting value is known as the half-wave potential ($E_{\frac{1}{2}}$). For a given supporting electrolyte, this potential is characteristic of the ion that is being reduced and is independent of the concentration of the ion. Half-wave-potential data are therefore of considerable value in qualitative polarographic analysis.

THE PRODUCTION AND INTERPRETATION OF POLAROGRAMS

The basic equipment for plotting curves relating current and voltage is simple; it consists of a variable potentiometer to apply the polarizing voltage to the electrodes, and either a galvanometer or a sensitive microammeter for measuring the current. The basic circuit is shown in figure 1. For extensive experimental work, however, equipment that is more automatic is preferable; a motor-driven potentiometer is used, and the current is recorded photographically or by a pen recorder. A typical current/voltage curve produced by such an instrument is shown in figure 3;

this is a polarogram for a solution consisting of 1 mg of a lead salt dissolved in 10 ml of 0·1N potassium chloride. It will be seen that a very small current, the residual current, flows through the cell before the reduction potential for the lead ions is reached. This current is associated with the formation of an ionic double layer at the interface between the dropping-mercury electrode and the solution. There is no way of eliminating this background current, and a correction for it is applied graphically, as indicated in figure 3. The polarogram also shows the small variations in the recorded current caused by the growth and fall of the mercury drops.

SUPPORTING ELECTROLYTE

The choice of a supporting electrolyte depends on several factors, the most important of which is the reduction potential for the cation of this electrolyte. This potential must be appreciably more negative than the half-wave potentials of the ions which are to be determined. Potassium, sodium, and ammonium salts are often used in the determination of the baser metals; tetramethyl-ammonium salts may be used for the determination of sodium and potassium. The anion of the supporting electrolyte is also important, as the polarographic method is not completely satisfactory for the study of reductions having half-wave potentials close to, or more positive than, the potential at which anodic dissolution of the mercury electrode takes place.

When two or more ions are to be determined in the same solution, the supporting electrolyte must be carefully chosen so that the $E_{\frac{1}{2}}$-values for the various waves are separated by at least 150 millivolts, otherwise the waves tend to coalesce and the heights of the individual waves cannot be accurately determined.

Frequently the diffusion current does not rise smoothly to its limiting value in the manner indicated in figures 2 and 3 and a sharp peak is observed at the beginning of the current-plateau. Such a maximum on a polarographic wave seriously interferes with the measurement of the wave height: maxima are usually suppressed by adding a small amount of a surface-active substance to the solution.

KINETIC AND CATALYTIC CURRENTS

The discussion has been limited, so far, to diffusion currents arising from simple electrochemical reductions. In some cases, however, it is possible for a purely chemical process, either preceding or following the electrochemical reaction, to affect the polarogram. Such a situation occurs in the reduction of certain aldoses and organic acids. With the aldoses, a tautomeric transformation of the ring form into an open chain precedes the electrochemical reduction, and the diffusion current is governed by the kinetics of the chemical transformation. Such transformations can be detected only by polarographic methods. Similarly, in the reduction of certain organic acids—a good example is pyruvic acid—only the undissociated form of the acid is readily reduced; again, the current is controlled by the rate of a chemical process.

Other reactions can give rise to catalytic currents. An example is the iron-catalysed hydrogen peroxide current. The reduction of ferric ions at the dropping-mercury electrode leads to the formation of ferrous ions. If hydrogen peroxide is present in the same solution, these ferrous ions are re-oxidized chemically to ferric ions while still in the neighbourhood of the dropping-mercury electrode, and they are thus again available for reduction at this electrode. The additional current caused by the presence of hydrogen peroxide is referred to as the catalytic current, and studies of this current have led to values for the velocity constant of the chemical reaction in fair agreement with values obtained by other methods. Another type of catalytic current is given by proteins when present in an ammoniacal solution. The current in these cases is due to the evolution of hydrogen, a process catalysed by the protein in solution.

APPLICATIONS OF CONVENTIONAL POLAROGRAPHY

Although there are many fields of application for the technique, its main use is in analytical chemistry. The limits of concentration for analytical determinations are from 10^{-2} to 10^{-5} M, application to lower concentrations being prevented by the occurrence of the residual current. It is possible to measure concentrations in the range 10^{-2} to 10^{-4} M with an accuracy of about ±2 per cent; the polarographic method is very suitable for trace-element analysis, for which the accuracy and reproducibility of the results are adequate. Moreover, determinations can often be completed in a fraction of the time needed for chemical methods, as there is, in many cases, no need for any lengthy chemical pretreatment of the sample. Human factors play a smaller part in polarography than in most analytical methods, especially when polarograms are recorded by automatic means. Such polarograms can be similarly interpreted by different workers, and the results can be checked at any time, as a permanent record is obtained. The advantage of this in routine analysis is readily apparent.

METALLURGICAL AND INORGANIC ANALYSIS

The uses of polarography in metallurgical analysis are very extensive and cover the determinations of trace amounts of Cu, Pb, Sn, Cd, Ni, Zn, and other elements in a wide range of ferrous and non-ferrous materials. In many cases it is necessary only to dissolve the sample in a mineral acid, but in some instances some means of preventing interference from a readily reduced major constituent of the sample must be employed. The simpler techniques for removing such interference involve changing the valency of the interfering element or the addition of a reagent to form a stable complex with the interfering species. Complex formation moves the polarographic waves to more negative potentials, and the value of this technique is exemplified by the analysis of copper alloys for lead and nickel. These alloys can easily be dissolved in hydrochloric acid, but unfortunately the polarogram of the solution shows a large wave due to the reduction of copper ions before those for the reduction of the ions of lead and nickel, which are present in brass. However, the determination of the lead and nickel contents can be completed polarographically without any further difficulty if the sample solution is made alkaline, and cyanide added to form complexes with the copper and zinc ions.

The uses of polarography in inorganic analysis are very numerous and include the determination of various inorganic constituents in ores, minerals, rocks, waters, inorganic chemicals, organic preparations, paint pigments, propellants, dyestuffs for use in foods, and such biological materials as plants and soils. Further details of the methods used in these branches of analysis are given in treatises on polarography [1, 2].

The polarographic method is not limited to ions that undergo reduction or oxidation at the dropping-mercury electrode. Elements whose ions are neither reducible nor oxidizable include thorium, zirconium, hafnium, and tantalum, but indirect methods are available for the determination of at least two of these elements: both thorium and zirconium can be precipitated with *m*-nitrobenzoic acid, and, after filtration and re-solution of the precipitate in mineral acids, the metal can be determined indirectly from the polarographic wave of the nitro-groups in the precipitate.

ORGANIC ANALYSIS

Polarography is not as widely applicable to organic analysis as to inorganic analysis. Numerous organic substances are reducible at the dropping-mercury electrode, but many of these reductions are irreversible processes giving rise to polarographic waves that approach their maximum heights rather slowly. Such waves are not generally as readily evaluated as those encountered in inorganic polarography. Moreover, hydrogen ions participate in the reduction of many organic substances, and it is necessary therefore to use solutions in which the hydrogen-ion concentration is carefully controlled. Another important consideration is the low solubility of many organic materials in aqueous solutions: analysis must sometimes be carried out with mixtures of organic solvents and water, or with non-aqueous solutions. Alcohol-water mixtures frequently give satisfactory results, and, among the non-aqueous solvents, ethylene glycol, dioxan, pyridine, and alcohol have proved useful. Organic compounds that undergo reduction at the dropping-mercury electrode generally contain reducible groups or conjugated double bonds, and useful waves are given by aldehydes, ketones, aliphatic and aromatic nitro-compounds, azo-compounds, vitamins, steroids, sulphur compounds, and organometallic substances [1, 2].

GAS ANALYSIS

Oxygen undergoes reduction at the dropping-mercury electrode to give two waves of equal height. The first of these is suitable for the determination of oxygen in various materials, including liquids, gases, and solids. The continuous determination of the concentration of dissolved oxygen in natural waters is possible when a dropping-mercury electrode with a wide bore is used. Polarography can also be used to study the oxygen content of blood plasma, and to investigate respiration and photosynthesis in living cells, such as in green algae.

USE IN TITRATIONS

The dropping-mercury electrode may be employed to determine the end point in a chemical titration. A typical example is the titration of lead with dichromate ions. The diffusion current given by an applied potential difference of 1·0 V will fall as the dichromate solution is added, until the end point is reached; after this, it will rise

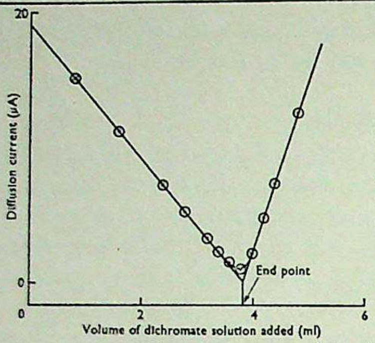


FIGURE 4 – *Titration curve of lead nitrate solution with potassium dichromate solution.*

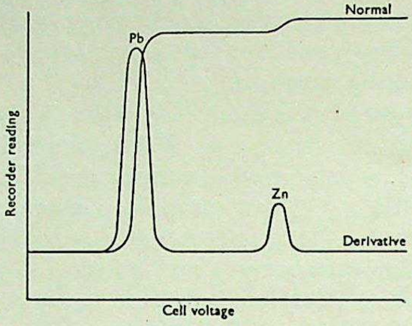


FIGURE 5 – *Normal and derivative polarograms (idealized) for a solution containing zinc and lead ions.*

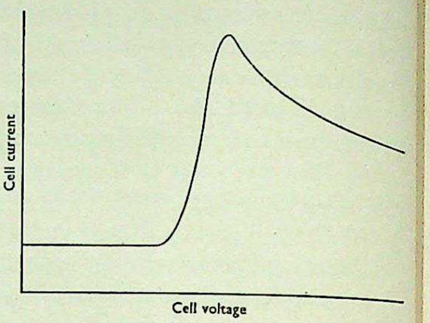


FIGURE 6 – *Current/voltage curve in oscillographic polarography.*

again. An extrapolation of lines drawn through the experimental points to their intersection gives the position of the end point (figure 4). This technique is also applicable when the species to be determined is reducible and the titrant is not reducible, and vice-versa. The break in the current/volume curve is not always sharp, because of the solubility of the precipitate, but this usually causes little difficulty in fixing the end point. This technique is used mainly for precipitation reactions, although there are some titrations based on complex ion formation.

DERIVATIVE POLAROGRAPHY

Attempts have been made, during the last ten years, to increase the resolving power of the polarograph by recording the rate of change of the current flowing in the cell as the applied voltage increases. These 'derivative polarograms' will show a maximum at the half-wave potential. The technique is particularly useful in the determination of a reducible ion in the presence of a larger concentration of a more readily reducible ion (figure 5).

Derivative polarograms may be obtained by immersing two synchronized dropping-mercury electrodes in the sample solution. The polarizing circuit for these electrodes is so arranged that the potential of one electrode lags behind that of the other by a small, constant, amount, and the current-recording circuit is modified to measure the difference in the diffusion currents at the two electrodes. Several variations on this method have been described [3, 4], but an alternative method which can be used with a normal polarographic cell has been more widely adopted because of the difficulties arising from the use of synchronized electrodes [5, 6]. Basically, this second method involves merely the insertion of some type of differentiating circuit in one of the electrical connections to the polarographic cell. The differentiating circuit may be a large condenser connected in series with the galvanometer or recorder, or it may be a relatively complex electronic circuit.

A.C. POLAROGRAPHY

Methods such as those mentioned above cannot increase the sensitivity of the conventional polarograph, and several new types of polarograph have been developed in recent years to meet a growing need for increased sensitivity combined with high resolution. These instruments produce derivative polarograms recording the alternating component of the cell current when a small alternating voltage is superimposed on the slowly changing voltage applied to the electrodes of the polarographic cell. The simplest of these instruments utilizes a small, sinusoidal, voltage, and the periodic component of the cell current then is the sum of two sinusoidal currents, one of which is connected with the ionic double-layer at the surface of the dropping-mercury electrode and the other with variations in the diffusion current produced by the small rapid changes in the cell voltage. The difference between the phases of these currents can be used to obtain a polarogram which shows principally the variation of the alternating component of the diffusion current with the average value of the cell voltage. This polarogram has much the same general shape as a derivative polarogram obtained by simpler methods, but reducible ions can be detected at an appreciably lower concentration than with the simpler instrument.

Even higher sensitivity is obtainable if a square-wave voltage is used in place of the sinusoidal voltage, as the former type of polarizing voltage permits a better separation of the periodic component of the diffusion current from the much larger current arising from the ionic double-layer. With a square-wave polarograph, reducible ions can be detected when their concentration is as small as 4×10^{-8} M, and high sensitivity is obtained even in the presence of large amounts of ions that interfere with the operation of the conventional polarograph [7]. For example, trace constituents in steels can be determined polarographically when the sample solutions contain gross amounts of readily reduced ferric ions. Many important applications for the square-wave polarograph have already been found in the analysis of such materials as alloys, natural waters, trade effluents, foodstuffs, fertilizers, and minerals [8].

The pulse polarograph is yet another of the electronic instruments developed in recent years to give both high sensitivity and high resolution [9]. This instrument applies small, rectangular-wave, voltage pulses ($\frac{1}{25}$ sec duration) to the electrodes of the polarographic cell. Each pulse occurs at a predetermined time in the life of the mercury drop, and the pulses are superimposed on a slowly changing voltage such as is employed in conventional polarography. Electronic circuits are used to measure the changes in diffusion current produced by the pulses. This instrument is appreciably more sensitive than the square-wave polarograph, especially for the determination of irreversibly reduced substances such as oxygen, hydrogen peroxide, and many organic compounds. There are many applications for this instrument in inorganic and organic analysis.

OSCILLOGRAPHIC POLAROGRAPHY

One further important development in recent years is a polarograph which displays a polarogram of a rather unusual form on the screen of a cathode-ray tube [10]. This polarograph applies to the electrodes of the cell a rapidly changing voltage that starts to change at a definite moment in the later part of the life of the mercury drop. The whole or part of the useful range of potential of the dropping-mercury electrode can thus be explored within the life of a single drop, and a polarogram is displayed almost continuously on the screen of the cathode-ray tube. When this type of instrument is used, the diffusion current produced by a simple reduction rises rapidly to a maximum value at a potential near to the normal half-wave potential and then decays slowly as the potential of the cathode becomes still more negative (figure 6).

Oscillographic polarography permits determinations to be carried out with great rapidity and with about the same accuracy as is obtained in conventional polarography. High resolution is obtained, and interference from a readily reducible major constituent can be largely avoided by making the potential of the dropping-mercury electrode before the start of the voltage sweep somewhat more negative than the half-wave potential for the reduction of the interfering ion. In favourable circumstances, reducible ions can be determined at concentrations down to about 5×10^{-7} M. This type of instrument, the single-sweep oscillographic polarograph, is used for routine analysis in many laboratories.

This instrument, and other modern instruments, make extensive use of electronic circuits to obtain improved performance and, as yet, the reliability of these instruments compares unfavourably with that of the conventional polarograph. When this defect is remedied, as it can be by using recent developments in the field of semiconductors, these instruments should play an important part in the further development of polarography as an analytical method. It also seems likely that some of them will be used to cast light on such fundamental problems as the mechanisms of electrode processes at mercury and other electrodes.

REFERENCES

[1] MILNER, G. W. C. 'The Principles and Applications of Polarography'. Longmans, Green & Co., London. 1957.

[2] KOLTHOFF, I. M. and LINGANE, J. J. 'Polarography' (second edition). Interscience Publishers, New York. 1952.

[3] HEYROVSKY, J. *Analyt. chim. acta*, **2**, 537, 1948.

[4] AIREY, L. and SMALES, A. A. *Analyst*, **75**, 287, 1950.

[5] VOGEL, J. and RIHA, J. *J. Chim. phys.*, **47**, 5, 1950.

[6] LEVEQUE, M. P. and ROTH, F. *Ibid.*, **46**, 480, 1949.

[7] BARKER, G. C. *Analyt. chim. acta*, **18**, 118, 1958.

[8] MILNER, G. W. C. *Proc. Congr. Mod. Analyt. Chem.*, St. Andrews, 1957.

[9] BARKER, G. C. and GARDNER, A. W. A.E.R.E. *Rep.* C/R 2297, 1958.

[10] DAVIS, H. M. and SEABORN, J. E. *Electron. Engng*, **25**, 314, 1953.

Problems of growth and calcium deposition in reef corals

T. GOREAU

The corals are a group of animals closely related to the sea anemones; they differ from them in being able to form massive calcareous skeletons that eventually become large, dense colonies. This article describes recent work using radioactive isotopes to study the mechanism of growth and calcium deposition in these colonies, and the formation of coral reefs.

Coral reefs form the most extensive marine communities known on this planet at the present time. Some of them have been in continuous existence for over 50 million years. Because corals form massive skeletons of calcium carbonate, the long-continued deposition of this has resulted in the accumulation of reefs, some of which are several thousand feet thick. As such, the reef corals are important land builders in tropical areas, frequently forming whole new islands and strikingly altering the shorelines of continents. The study of reefs is also of special interest because limestone formation takes place here under much the same conditions as must have occurred during earlier times in the earth's history. The simple methods of descriptive natural history and field geology, such as those used in the 1830s by Charles Darwin and James Dwight Dana in their classic work on the origins of coral islands, are today supplemented by a variety of new techniques, ranging from free diving to mass spectrometry and micropalaeontology.

The biological study of corals began with the taxonomic investigations of Lamarck and Milne-Edwards, among others, in the early nineteenth century. The principal features of the anatomy, histology, and embryology of the corals were worked out by the early part of this century; such investigations continue today using the new techniques of electron microscopy and chemical cytology. Towards the end of the nineteenth century, Alexander Agassiz started his important work on the growth rates and ecology of reef corals. Similar studies were later carried out on a world-wide scale by others [1], thus establishing the significance of coral reefs in relation to major changes in land and sea-level, especially during the Pleistocene geological period. Recent work by ecologists has also brought to light the amazing ability of coral-reef communities to maintain a very high productivity and standing crop of organic matter, despite the fact that most reefs occur in the impoverished surface waters of tropical seas.

The experimental investigation of the reef-building corals as such started only about fifty years ago. The physiological studies carried out on nutrition and growth of corals by members of the British Great Barrier Reef Expedition of 1928–29 are an outstanding example. Important contributions to our knowledge of reef-coral physiology and biochemistry were also made by Japanese marine biologists working in Palau before World War II. Nevertheless, there are still major problems of fundamental importance to be solved, and it is the object of this paper to deal with some of these within the framework of our own recent studies on reef-coral skeletogenesis.

The true reef-building corals are limited in geographical distribution to the clear warm sunlit waters of tropical oceans; they centre upon the great reef-tracts of the Indo-Pacific and Western Atlantic coral seas. The unit of structure in most reefs is the coral polyp. However, the building activities of the corals are often enhanced by the calcareous algae, which cement and consolidate the coral framework into a solid mass of limestone.

The so-called stony corals, the Madreporaria or Scleractinia, are an order of animals belonging to the phylum Coelenterata, class Anthozoa, subclass Hexacorallia [2]. Their most striking feature is their ability to form a massive, calcareous skeleton, which forms a supporting base upon which the polyps live (figure 1). Although an individual polyp is only of the order of 10 mm in diameter, such skeletal masses often grow to enormous sizes as a result of colony formation by the asexual division of mature polyps. Individual coral-colonies weighing several hundred tons and large enough to fill a large living-room are common on many reefs. The closest relatives of the true corals are the non-calcareous sea anemones, familiar to anyone

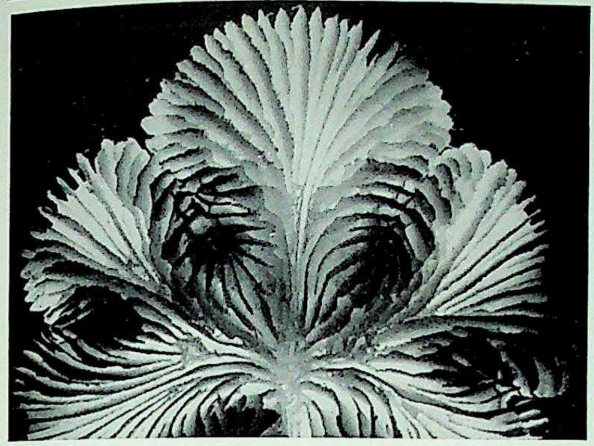


FIGURE 1 – *Skeleton of* Colpophyllia natans, *a reef-building coral. A young colony is shown; the fleshy parts have been removed. The radial lines are the calcareous ribs.*

who visits the seashore. The corals resemble the anemones in basic body structure and overall appearance; they differ in that their basal disk is gossamer-thin and highly folded, offering an enormous surface area for limestone secretion.

The terms growth and skeletogenesis, in this paper, refer to the calcification of corals. The terms polyp and corallite are used interchangeably here to describe the individual animal unit of a coral colony. The word 'coral' is itself confusing, as it is used by laymen to describe practically any branching calcareous or horny growth of marine origin. In this paper the term coral refers specifically to Madreporaria or Scleractinia, but not, for example, to seafans or the precious 'red coral' that belongs to the Octocorallia, or the 'stinging coral', which is a calcareous hydrozoan. In the corals, the calcareous skeleton is called the corallum or sclerenchyme; the fleshy parts are collectively known as the polyparium or the coenenchyme. Sharp distinctions are usually made in the technical literature between the deep-water, non-reef-building corals and the tropical, shallow-water, reef-building corals.

A remarkable feature of all reef-building corals is their symbiosis with unicellular algae known as zooxanthellae. Tropical shallow water corals contain large numbers of these within specialized carrier cells in the gastrodermal epithelium. The zooxanthellae are Dinophyceae, a group to which many of the free living marine flagellate algae also belong. An electron micrograph of zooxanthellae is shown in figure 8.

TECHNIQUES

There is one major obstacle to the study of the growth of corals. This is the difficulty of keeping corals alive and healthy in laboratory aquaria for sufficient time for accurate growth-measurements to be made. Such experiments take weeks or even months. We wished to devise a more rapid method, and we started four years ago to develop isotope-tracer techniques using radioactive calcium-45. Corals are unable to distinguish between the stable and radioactive varieties of calcium; both isotopes are deposited into the skeleton in the same ratio as they are present in the medium. The amount of radioactivity taken up under various controlled experimental conditions is therefore an index of growth and physiological function [3]. This technique is so sensitive that growth can be measured in corals that have been exposed to calcium-45 for only a few hours, and it can be adapted to the study of coral growth in the actual

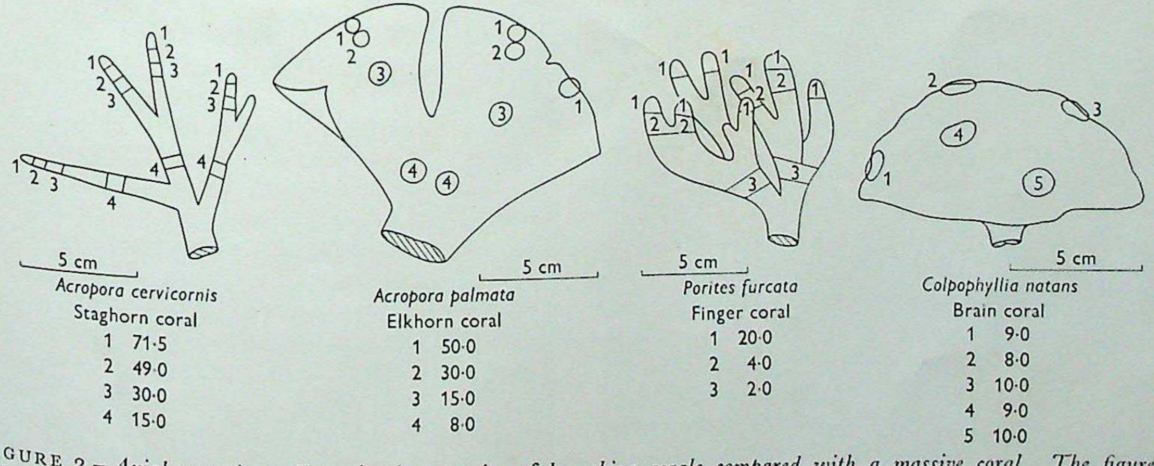


FIGURE 2 – *Axial growth gradients in three species of branching corals compared with a massive coral. The figures below each specimen are the average growth-rates, given in µg Ca per mg of tissue nitrogen per hour, in the different parts of the colonies that are identified by numbers.*

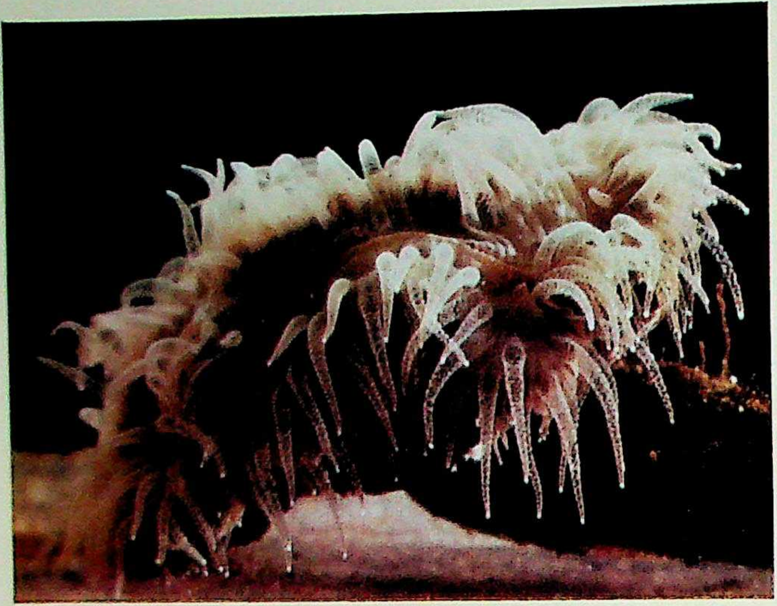


FIGURE 3 – Meandrina braziliensis, *a free-living coral. This young colony is shown in the feeding attitude, and is completely expanded, with a large number of tentacles bearing batteries of nematocysts, which can be seen as small white spots.*

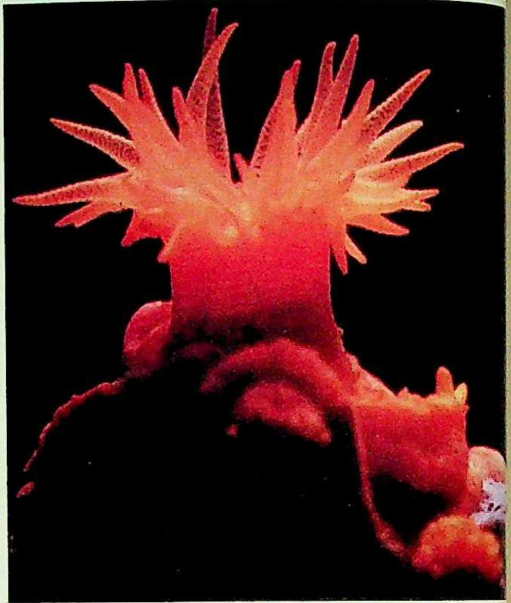


FIGURE 4 – Tubastrea tenuilamellosa, *a typic non-reef-building coral. Unlike the other species show here, this coral has no zooxanthellae.*

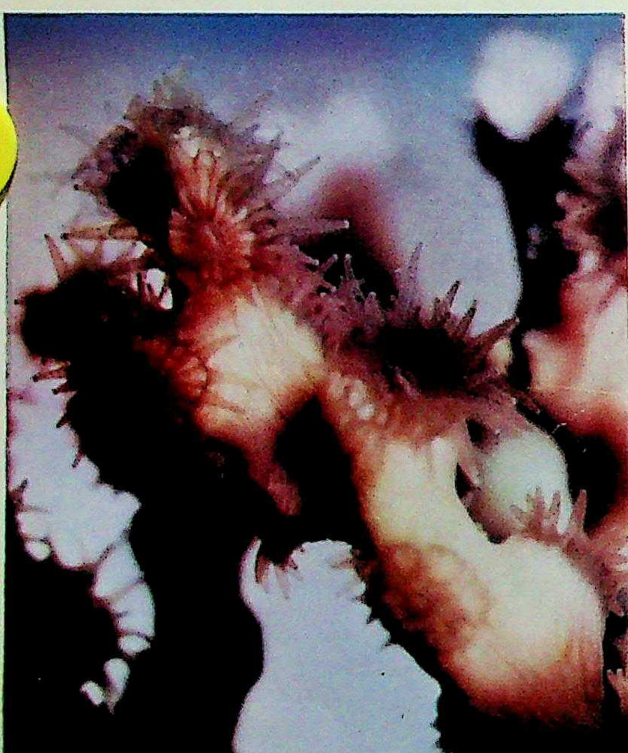


FIGURE 5 – Oculina diffusa, *a small branching coral. The colour is due to the zooxanthellae. The corallites are about* 5 *mm in diameter.*

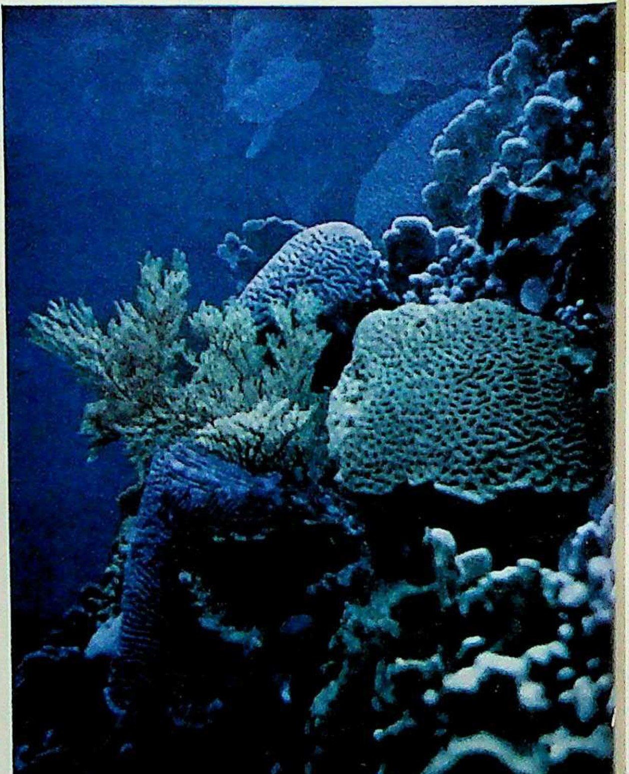


FIGURE 6 – *Underwater photograph showing the side of a co buttress at a depth of 20 ft. The large round masses in the cen are reef-building scleractinian corals. The plate-like forms in upper right-hand corner are colonies of* Millepora, *the pois 'coral', a reef-building hydrozoan.*

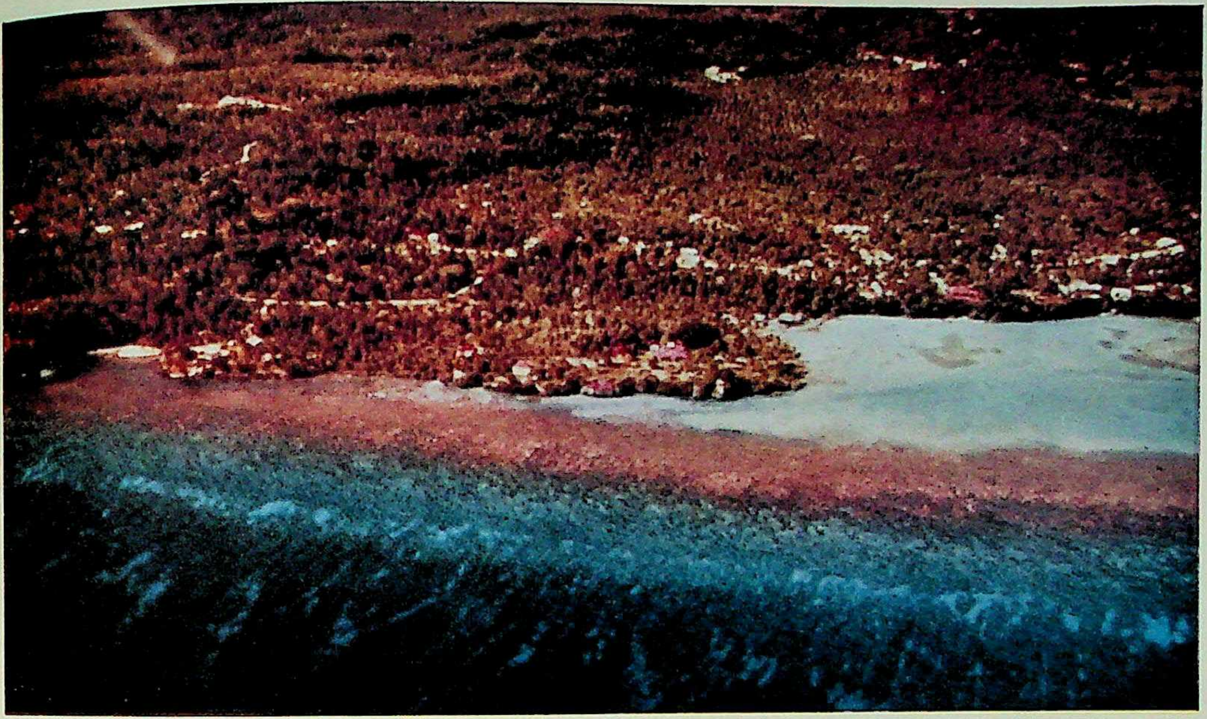


FIGURE 7 – *Aerial view of a fringing reef near Ocho Rios, Jamaica, showing marked structural zoning. The reef is separated from the shore by a light-blue shallow lagoon. The distance across the middle of the photograph is about 700 yards.*

reef [4]. As only small samples are needed for analysis, it is possible to measure differences in the growth rate in various parts of a single colony or even within a single corallite. Although many basic questions of coral skeletogenesis remain to be answered, the use of these new techniques has at least made it possible to describe some of the factors that control and influence the calcification reaction.

FACTORS REGULATING GROWTH RATE OF CORALS

The most important of the intrinsic factors influencing calcification in corals are the species; position in the colony; the age of the colony; and enzymes.

Although reef-building corals grow in fairly uniform conditions of temperature, water circulation, and illumination, we have consistently found very large differences in the growth rates of different species. The highest rates are invariably found in the branching, acroporid corals (the West Indian elkhorn and staghorn corals), with Millepora (a hydrocoral-line) a close second, and the Poritidae (finger corals) third. The various massive corals are usually slower [4].

The position in the colony affects the growth rate in the branching corals, but not in the massive corals. In the former we found that most of the growth took place in the terminal corallites at the tips of the branches; in the latter the rate was more or less the same all over [4] (figure 2). The factors that determine the polarity of growth are not understood. In the staghorn corals, new branches may arise almost anywhere on the older parts of the colony. Their appearance is heralded by the differentiation of large pale apical corallites; these are very different from the much more numerous small dark polyps that form the sides of the branches. In the finger corals, on the other hand, new branches arise mostly through the dichotomous growth of older branches, and there is little obvious differentiation of the apical polyps, save for their much greater growth-rate.

The effect of age on growth-rate is considerable in some corals, negligible in others. T. Stephenson and A. Stephenson [5] showed that the growth-rate of some corals decreased in inverse proportion to their size and age. In the West Indian rose coral (*Manicina areolata*) we found that colonies weighing approximately 0·05 g deposited calcium nearly one hundred times faster, per unit of tissue nitrogen, than did colonies weighing 150 g

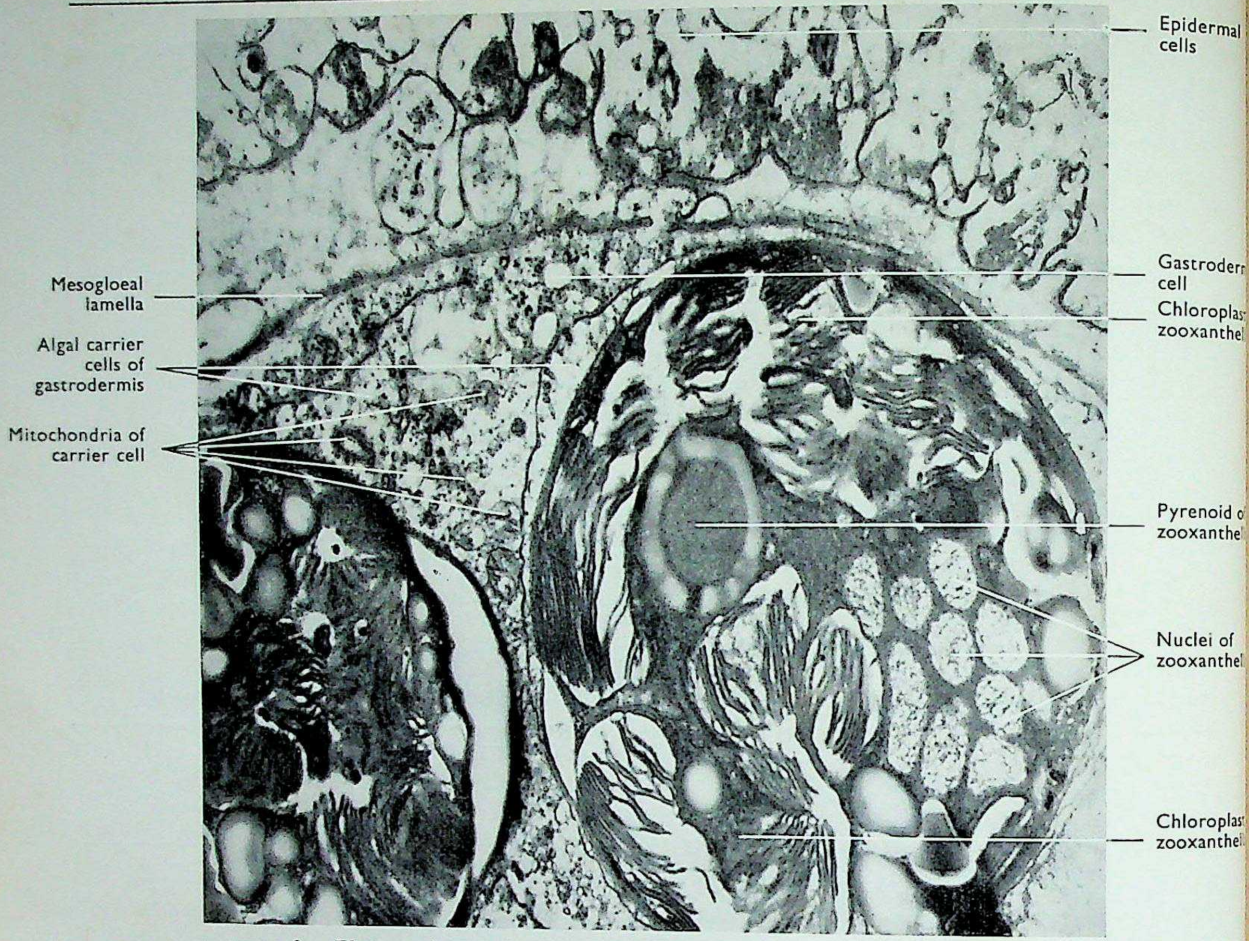


FIGURE 8 – *Electron micrograph of two zooxanthellae in the gastroderm of a coral. Magnified* 17 400 *times. The zooxanthellae are contained within specialized carrier-cells, which can be seen as protoplasmic rings surrounding the algal cells. Note the mitochondria in the cytoplasm. Each zooxanthella contains a nucleus, pyrenoid, and chloroplast, the last of which may be seen as the large striated horseshoe-shaped object that nearly fills the algal cell.*

[6] (figure 9). We believe that this phenomenon may be connected with the peculiar mode of life of the coral. *M. areolata* is usually found unattached, and on sandy or soft grassy bottoms where a heavy skeletal mass would result in the colony's sinking into the sediment. The slowing of the growth-rate with increasing size may thus be an adaptive mechanism for limiting the size of the species in response to the necessity of keeping 'afloat' on top of shifting, unstable, bottom-sediments. We do not believe it likely that there is much slowing down of growth with age or size in those hermatypic corals that habitually reach great sizes on the reef, or in branching species in which colony growth is indeterminate, resulting in the formation of irregular underwater coral-hedges, often many acres in extent.

We have so far investigated two enzymes to see if they are connected with calcification in the corals. The finding of alkaline phosphomonoesterases in corals [7] led us to propose that phosphatases were involved in calcification; subsequent histochemical studies showed that these enzymes were absent from any structures associated with skeletogenesis [8]. In a series of unpublished experiments we demonstrated that most coelenterates, including all twenty-four species of coral tested, contained high concentrations of carbonic anhydrase. This enzyme catalyses the reversible hydration of CO_2 to H_2CO_3 and also the dissociation of carbonic acid to carbonate. Carbonic anhydrase has been shown to be involved in carbonate shell deposition in the chicken egg [9], the oyster [10], and the barnacle [11]; we therefore investigated the possibility that it also plays a role in coral skeletogenesis. The use of low-toxicity specific carbonic-anhydrase inhibitors brought about a remarkable reduction in the calcium-deposition rate in corals. When the corals contained zooxanthellae, the reduction of the growth-rate was far less in the light than in the darkness; if zooxanthellae were absent, the degree of inhibition was the same in light as in darkness, and the growth-rate fell to extremely low levels [3]. A discussion of the significance of these findings will be deferred until later.

Light appears to be the most important of all the environmental factors that control the calcification rate of hermatypic corals in the reef. It has long been known that reef-building corals can grow well only in shallow water where the light is intense enough to allow a net excess of photosynthetic production of organic matter by plants. Reef corals always grow towards the light, and they are hardly ever found in caves or other dark places where there is too little light for photosynthesis in the zooxanthellae.

We designed experiments to test the effect of variations in light intensity upon the rate of calcification of reef-building corals under approximately natural conditions; we found that calcium uptake is fastest at noon on a clear sunny day; is reduced by approximately 50 per cent on a cloudy day; and is reduced by an average of nearly 90 per cent in total darkness [4]. The light : dark ratio in the calcification rates of different species varied between approximately 3 : 1 and 23 : 1, averaging 10 : 1. This striking dependence of coral growth-rates on ambient-light intensity is dependent on the presence of the zooxanthellae; this is shown by the extremely low calcium-deposition rates of otherwise healthy reef corals from which the zooxanthellae have been removed by culturing in total darkness for several months before the experiment. The calcification rate of such colonies was very low and was independent of the light intensity [3].

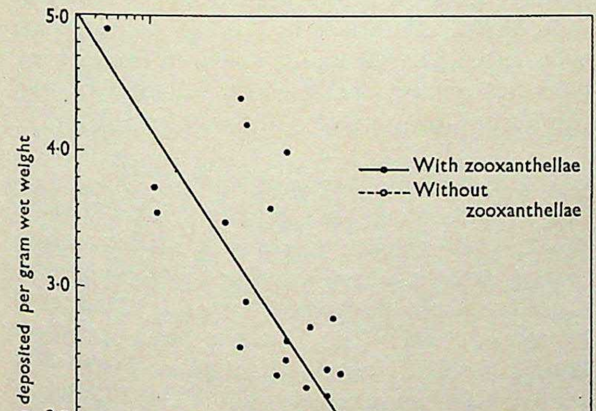


FIGURE 9 – *Calcification rate plotted against size in* Manicina areolata, *a small free-living brain coral. The data show that the smallest colonies grow approximately* 100 *times faster per unit weight of coral than the largest.*

THE COMPOSITION OF THE CORALLUM

An important, and still largely unsolved, question of coral calcification concerns the origin of the calcium deposited by the corals. Since most of our experiments were carried out on non-feeding corals exposed to radioactive calcium added to the water as $^{45}CaCl_2$, it is fairly clear that the corals must be able to absorb Ca^{++} directly from seawater, and that very little is derived from their food, which consists of small planktonic animals. Indirect support for this view is provided by evidence that corals are apparently unable to discriminate calcium from strontium, so that both elements are deposited in the skeleton in the same proportions as they exist in seawater [12]. Most organisms, especially the crustacea that form a large part of the food supply of corals, appear to exclude strontium; the coral shows no sign of this discrimination as it would be expected to do if its calcium were derived from its food.

There is no definite information about the mechanisms responsible for transport of the calcium inside the coral. Calcium does not appear to be concentrated in the tissues before deposition in the skeleton; on the contrary, corals tend to have a lower concentration of calcium in tissues than there is in seawater [13]. As $CaCO_3$ is slightly soluble in seawater, there is a considerable amount of equilibrium exchange of calcium between the corallum and the medium; no net solution can occur under normal conditions, because the seawater in coral reefs is supersaturated with respect to $CaCO_3$. The tendency of the corallum to exchange calcium with the seawater is drastically reduced as long as it is covered by a living, intact sheet of polyps. This protective action disappears as soon as the polyparium is killed; the

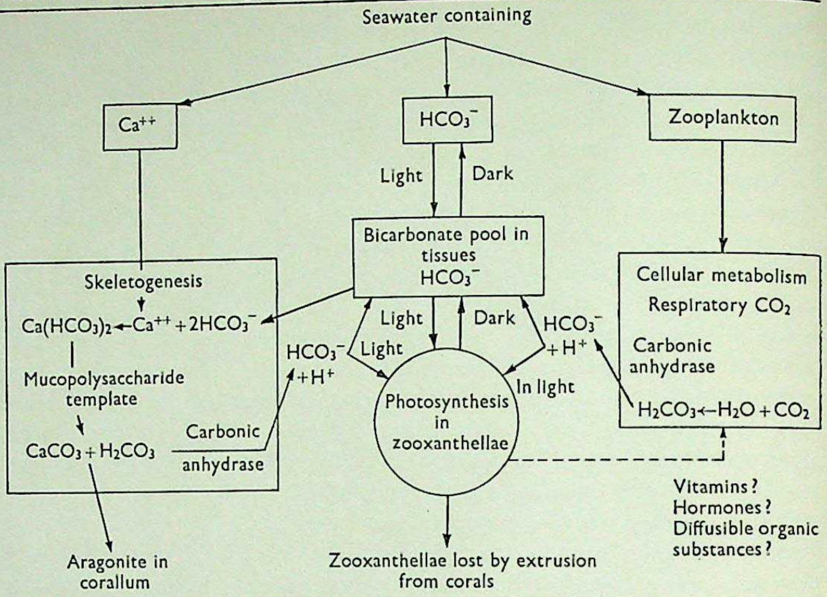

Seawater containing
Ca^{++}
HCO_3^-
Zooplankton
Light
Dark
Bicarbonate pool in tissues HCO_3^-
Skeletogenesis
$Ca(HCO_3)_2 \leftarrow Ca^{++} + 2HCO_3^-$
Mucopolysaccharide template
$CaCO_3 + H_2CO_3$
Carbonic anhydrase
HCO_3^- $+H^+$
Light
Dark
Light
HCO_3^- $+H^+$
In light
Photosynthesis in zooxanthellae
Cellular metabolism
Respiratory CO_2
Carbonic anhydrase
$H_2CO_3 \leftarrow H_2O + CO_2$
Vitamins?
Hormones?
Diffusible organic substances?
Aragonite in corallum
Zooxanthellae lost by extrusion from corals

FIGURE 10 – *Possible pathways of calcification in relation to tissue respiration and photosynthesis by zooxanthellae in reef corals.*

equilibrium exchange rate then increases by several orders of magnitude. This shows that the polypary layer is fairly 'calcium proof', and that it effectively isolates the skeleton from the environment as long as it is alive [14]. This is the opposite of what is seen in vertebrates, where the calcium of the bone remains relatively mobile and freely exchangeable with the calcium dissolved in the body fluids.

The coral skeleton is also remarkable for being composed entirely of $CaCO_3$ with the crystalline structure of aragonite; calcite is completely absent. This probably has some important consequences with regard to the chemical composition of the corallum and the geographic distribution of the corals as a group. Since the skeleton is pure aragonite, magnesium carbonate is almost absent from the corallum, despite the fact that magnesium carbonate is more abundant in seawater than is calcium carbonate. Those calcareous organisms that deposit calcite as well as aragonite have at least ten times more $MgCO_3$ in their skeletons than the corals have [15]. H. Lowenstam has suggested [16] that the failure of corals to produce calcite may be the factor responsible for the marked reduction in the number of scleractinian species in cold waters compared with their abundance in warm waters. Organisms capable of producing both calcite and aragonite tend to deposit more of the former during cold seasons, and more of the latter during warm seasons. Since corals cannot substitute the less soluble calcite for aragonite in cold water, their calcification mechanism may become so inefficient as to hinder skeletogenesis under these conditions.

The application of tracer methods using stable isotopes has led to some important advances in our knowledge regarding the probable origin of the carbonate fraction of the coral limestone. It has been pointed out by a number of investigators [17–19] that the $^{18}O : {}^{16}O$ and $^{13}C : {}^{12}C$ ratios of the carbonate deposited by reef corals are not the same as in the carbonate dissolved in seawater. C. Emiliani [19] has suggested that the oxygen-isotope ratio might be altered if chemical exchange occurs between the oxygen that is liberated by zooxanthellae and the carbonate that is deposited. If the skeletal carbonate of the corals is derived from metabolism rather than from the seawater, then, again, alterations in the isotopic ratios are to be expected. Metabolic reactions produce isotopic fractionation because of the slightly different equilibrium-constants for different nuclides.

THE MECHANISM OF CALCIFICATION

The isotopic evidence for a metabolic origin of the skeletal carbonate agrees with our own experimental data; it is now possible to define some of the more probable pathways in the calcification process in corals. This is only a first approximation, a working hypothesis that may also find general application to processes of $CaCO_3$ deposition in the other calcareous organisms.

It appears (figure 10) that calcium is absorbed from the seawater as Ca^{++} and transferred by an as yet unknown mechanism to the site of calcification, which is probably localized just outside the calicoblastic epidermis. At this point the Ca^{++} is absorbed on a mucopolysaccharide-like material that forms part of the organic matrix and acts as the template upon which the initial stages of skeleton mineralization occur. Here the Ca^{++} combines with bicarbonate of primarily metabolic origin to form calcium bicarbonate and then calcium carbonate. The rate of calcium carbonate precipitation would, in principle, be governed by the overall rate of the coupled system

$$Ca^{++} + 2HCO_3^- \rightleftharpoons Ca(HCO_3)_2 \qquad (1)$$

$$Ca(HCO_3)_2 \rightleftharpoons CaCO_3 + H_2CO_3 \qquad (2)$$

$$H_2CO_3 \overset{\text{carbonic anhydrase}}{\rightleftharpoons} H^+ + HCO_3^- \qquad (3)$$

$$H_2CO_3 \overset{\text{carbonic anhydrase}}{\rightleftharpoons} CO_2 + H_2O \qquad (4)$$

at the site of mineralization. A crucial factor is the rate at which H_2CO_3 is removed from the system by reactions 3 or 4, both of which are catalysed by carbonic anhydrase. As tropical marine surface-waters are supersaturated with respect to $CaCO_3$, small changes in the velocity of reactions 3 and 4 will greatly increase the velocity of reactions 1 and 2. Under these conditions the free-energy changes are very small and the work done by coral in skeletogenesis is consequently also small. If the scheme is essentially correct, it is easy to see how the presence of photosynthesizing zooxanthellae causes such a marked increase in the growth rate of reef corals. Both CO_2 and HCO_3^- are fixed by algal photosynthesis. This would have the same effect as carbonic anhydrase in increasing the overall reaction-rate of the system. When reef corals are exposed to light, the zooxanthellae and carbonic anhydrase exert their effects synergically; in darkness only the latter can act.

CONCLUSIONS

Although there can now be little doubt that zooxanthellae assist coral skeletogenesis by a mechanism similar to the one outlined above, these symbionts probably have other important effects on the bio-economy of corals that cannot be fully described here. Their intimate structural relation to the coral host is shown in the electron-micrograph in figure 8. There is still a lively controversy between those who hold that the only function of the zooxanthellae is to supply the coral with oxygen and remove carbon dioxide [20], and those who believe that zooxanthellae are a source of food as well as of oxygen [21]. The writer favours the view that the zooxanthellae increase the metabolic efficiency of the animal host both through *in situ* production of oxygen and absorption of inorganic waste products by photosynthesis, and through the secretion of trace amounts of vitamin-like or hormone-like factors [3, 22]. The writer agrees with C. M. Yonge [23] that the zooxanthellae do not appear to be food for corals.

Apart from the problem of trophic relationships between the scleractinian corals and their algal symbionts, it is quite certain from our experimental evidence that, through their chemical potentiation of the calcification process, the zooxanthellae must have played a major role in the evolution and growth of shallow-water coral-reefs. The development of such enormous communities in the face of constant battering by heavy seas was possible only when processes of limestone deposition became fast enough for the rate of accumulation to exceed the rate of loss by organic and inorganic attrition. The synergistic effect of the zooxanthellae on the calcification rate of corals provided the necessary stimulus and was one of the decisive factors in the evolution of coral reefs.

ACKNOWLEDGMENTS

Much of our work in Jamaica is part of a research programme that is being carried out jointly with the New York Zoological Society, with support of the National Science Foundation and the Office of Naval Research of the U.S. Navy.

REFERENCES

[1] Gardiner, J. 'Corals and Atolls'. Macmillan & Co. Ltd, London. 1931.
[2] Hyman, L. 'The Invertebrates'. Vol. 1, McGraw-Hill Book Co. Inc., New York. 1940.
[3] Goreau, T. *Biol. Bull., Woods Hole*, **116**, 59, 1959.
[4] Goreau, T. and Goreau, N. *Ibid.*, **117**, 239, 1959.
[5] Stephenson, T. and Stephenson, A. *Sci. Rep. Gr. Barrier Reef Exped.*, **3**, 167, 1933.
[6] Goreau, T. and Goreau, N. *Biol. Bull., Woods Hole*, **118**, 419, 1960.
[7] *Idem. Proc. Nat. Acad. Sci., Wash.*, **39**, 1291, 1953.
[8] *Idem. Nature, Lond.*, **177**, 1029, 1956.
[9] Benesch, R., Barron, N., and Mawson, C. *Ibid.*, **153**, 138, 1944.
[10] Wilbur, K. and Jodrey, L. *Biol. Bull., Woods Hole*, **108**, 359, 1955.
[11] Costlow, J. *Physiol. Zoöl.*, **32**, 177, 1959.
[12] Odum, H. *J. Inst. Mar. Sci. (Texas)*, **4**, 39, 1957.
[13] Goreau, T. and Bowen, V. *Science*, **122**, 1188, 1955.
[14] Goreau, T. and Goreau, N. *Biol. Bull., Woods Hole.* (In press.)
[15] Chave, K. *J. Geol.*, **62**, 267, 1954.
[16] Lowenstam, H. *Ibid.*, **62**, 284, 1954.
[17] Lowenstam, H. and Epstein, S. *Ibid.*, **65**, 364, 1957.
[18] Craig, H., quoted by Revelle, R. and Fairbridge, R., in *Mem. geol. Soc. Amer.*, **67**, (I), 239, 1957.
[19] Emiliani, C. *J. Geol.*, **63**, 538, 1955.
[20] Yonge, C. M. *Sci. Rep. Gr. Barrier Reef Exped.*, **1**, 353, 1940.
[21] Odum, H. and Odum, E. *Ecol. Monogr.*, **25**, 291, 1955.
[22] Goreau, T. and Goreau, N. *Science*, **131**, 668, 1960.
[23] Yonge, C. M. *Mem. geol. Soc. Amer.*, **67**, (1), 429, 1957.

The beginnings of phosphorus manufacture

Arthur Albright, who effectively started the British manufacture of phosphorus, was born 150 years ago. This is therefore an appropriate time at which to consider the history of the discovery and the early manufacture of this element.

Phosphorus was discovered by Hennig Brand, an alchemist and quack, in about 1670, when he was distilling human urine in an attempt to find a substance that would change silver into gold. (The carbon produced by the charring of the organic residue would act as a reducing agent for the phosphorus compounds normally present in urine.) The element—he called it *kaltes Feuer*—was a chemical curiosity, and Brand achieved a considerable social success as well as scientific fame. Johann Kunckel heard of the discovery and, together with his friend Johann Daniel Krafft, visited Brand. According to the generally accepted story, Brand would not say how he made phosphorus, and Kunckel had to rediscover the method. He also devised a method for casting phosphorus into sticks, and advertised it for medical use in his treatise on 'Phosphorus Mirabilis and its wonderful shining pills'. It was, however, Krafft who first demonstrated phosphorus in England, at the court of Charles II and also before the Royal Society. Boyle saw this demonstration, and wrote enthusiastically of the way that the 'twinkling sparks, without doing any harm to the Turkey Carpet they lay on, continued to shine for a good while'. Whether or not Boyle had again to rediscover the method of preparation is not certainly known; he himself stated that the only information that Krafft gave him was a hint stating that the 'principal matter' was 'somewhat that belonged to the body of man'. What is certain is that it was Boyle who first published the preparation of phosphorus as an aqueous solution or suspension, in his 'The Aerial Noctiluca' (1680), and as a solid, in 'The Icy Noctiluca' (1681–82); and he also first scientifically described the element and its properties.

Boyle's assistant, Ambrose Godfrey Hanckewitz, established himself as a pharmacist, under the name of Godfrey. He advertised the usual contemporary pharmaceuticals, such as smelling salts, Glauber's salt, balsams, and cordials, and also that he was 'the only one in London who makes inflammable phosphorus . . . All unadulterated, wholesale 50 shillings an ounce, and retail three pounds sterling an ounce'. Even at this price there was a ready market for phosphorus, and Godfrey travelled throughout Europe selling it, dying a rich man in 1741.

The high cost of phosphorus was due to the very poor yield of the process and remained high until Scheele, in 1777, devised a method of preparing it from bone-ash. The first commercial application of this process in Britain effectively dates from 1844, when Arthur Albright persuaded Edmund Sturge, with whom he had gone into partnership a few years earlier, to start making phosphorus. The firm was then at Selly Oak, Birmingham, but it shortly moved to Oldbury. The aim was to provide phosphorus for the British match trade; the native supplies at that time were very small indeed—about a ton a year—and a large amount had to be imported.

A French gelatine-making company, *La Société Vve. Dupasquier, fils et Coignet*, had started making phosphorus in 1838, and manufacture in Germany began at about the same time. Imports to Britain reached a peak value of over £2500 in 1844. When Albright decided to compete with the French firm, he had the disadvantage of having to import bones, but against this had what proved to be the decisive advantage of cheap coal. Black Country coal could be bought for one and sixpence a ton; it has been estimated that the French company was paying fifteen shillings a ton. Imports fell rapidly; by 1851 Britain was actually exporting phosphorus.

The usual raw material was bone-ash from South America, although Albright did make an adventurous trip to Galatz, on the Danube, in order to see whether he could use the residue from the slaughterhouses of the meat-canning industry that had been established there. It proved impossible to ship the bones before treatment, and he was forced to set up calcining furnaces on the spot. The bone-ash—in later years rock phosphates were used—was treated with sulphuric acid, and the resulting phosphoric acid was concentrated, mixed with charcoal, dried, and ground so as to be ready for the retorts. These retorts—bottle-shaped fireclay vessels about four feet long—were mounted in coal-fired furnaces in groups of twenty-four; they had iron necks that led the phosphorus vapour to the condensers. The distillation was continued for about sixteen hours; the phosphorus was condensed under water to prevent oxidation by the air, and was ladled off and allowed to set into thirty-pound cakes that were bleached with chromic acid. The bleached phosphorus was sent down a steam-jacketed cylinder to keep it molten, filtered through chamois leather, and cast into sticks.

Contemporary accounts give a vivid picture of British phosphorus-manufacture at that time. The phosphorus distillery of those days was a striking sight with 'long yellow flames' and with pieces of burning phosphorus 'spitting forth in fiery balls'; there was an incessant bubbling of the condensing phosphorus in the water basins, and an 'intolerable stench' of phosphorus derivatives. The combination of this factory with the adjacent Chance's soda-making factory, using the Leblanc process, with its vast output of offensive alkali-waste, must have contributed to turning Oldbury from what a local historian described as a pleasant and pretty place, nestling below the green slopes of Rowley Hills, into a waste, 'a town of unrelieved, almost defiant ugliness'.

Many improvements in manufacturing technique originated in this British factory. The wasteful directly heated furnaces were replaced by retorts fired by producer gas and using the Siemens regenerative principle; this principle was in fact used both here and in Chance's glass-works before it was widely adopted by steel manufacturers. These furnaces in turn gave way to electrical furnaces, which were continuous in operation; they depended on a process, that of preparing phosphorus by heating the bone-ash with silica (sand) and carbon, that had been proposed by Wöhler in 1829. The success of the process depends on attaining a temperature sufficiently high for the slag to be tapped off while molten. The first attempts to work it—by an Edinburgh man, J. B. Readman—used a blast furnace, but a sufficiently high temperature could not be attained with this. Blast furnaces were eventually used for phosphorus manufacture in the United States; even when the technical difficulties had been overcome it was found that they needed to be very large to be economically successful. Readman also took out a patent for a similar process using electrical heating, and this patent was taken up by the Electric Construction Company of Wolverhampton. This group formed the Phosphorus Company in 1890, and started operations using the electric furnace with two horizontal electrodes that is still frequently represented in textbooks of chemistry. This furnace is, in fact, obsolete, and was replaced by one of an improved design, using vertical electrodes, soon after the Albright company, by then Albright and Wilson, took over the firm.

Albright was deeply disturbed by the dangers to health and property that were involved in phosphorus manufacture. Yellow phosphorus, the allotrope prepared by all the processes so far mentioned and the one then used in matches, was both spontaneously inflammable and poisonous, causing a bone gangrene known as 'phossy jaw'. Anton Schrötter, of Vienna, had discovered in 1845 that a non-poisonous, stable, form, red phosphorus, could be made by heating yellow phosphorus in an inert atmosphere. The process could be explosive, as the reaction is exothermic, and he had decided that it had no commercial value.

In 1849 Schrötter was Albright's guest at a British Association meeting at which he read a paper on the new allotrope. Albright decided that the process was practicable, bought Schrötter's patent, and set to work to devise a way of avoiding explosions. He and Julius Lowe, one of his staff, designed a cast-iron pressure pot, having a small-bore tube as a safety valve, that eventually eliminated the risk of explosions. Albright then started a campaign to persuade European match manufacturers to use red phosphorus in the place of yellow.

It proved to be extremely difficult to make a satisfactory 'strike anywhere' match using red phosphorus: an alternative approach was to try to make matches using the 'safety-match' principle, in which all the phosphorus is on the box, none being incorporated in the head of the match. The first successful ones were made by the Lundström brothers of Sweden in 1855. They immediately ordered such a large quantity of red phosphorus that Albright imagined that it must be intended for making munitions and, as a Quaker, refused to supply them. He was, however, reassured, and sent the red phosphorus; the Lundström brothers started to manufacture matches on a large scale. This discovery of a satisfactory way of using red phosphorus for safety matches was the basis of the successes of both the Swedish match industry and Albright and Wilson, and it marked the beginning of the modern period of phosphorus manufacture and use.

Isotope separation by reversible chemical processes

E. GLUECKAUF

When the existence of isotopes was first recognized it was believed that, because of the identity of their electronic configurations, their separation was possible only by physical methods. The preparation of deuterium by electrolysis of water demonstrated that chemical methods of separation were also possible, and today chemical methods are not only widely used but are sometimes the method of choice. This article discusses the theoretical principles, and considers the advantages and limitations of chemical methods of isotope separation.

GENERAL CONSIDERATIONS

In recent years the uses of separated stable isotopes have become very varied. By far the largest number of all the separated products are used for studies of nuclear properties, but substantial use is made of them in the study of physico-chemical problems, for chemical analysis by the isotope-dilution technique, for medical and biological investigations, and as raw material for the preparation of radio-isotopes. But potentially valuable though they are, the difficulty of separating isotopes is a limiting factor in their use. To give a typical example, the radioactive isotope ^{47}Ca, with a half-life of 4·7 days, could, if readily available, be one of the most useful isotopes in the diagnosis of cancer. It can be produced by neutron irradiation of ^{46}Ca, a naturally occurring isotope; but though calcium is plentiful, the ^{46}Ca content of the natural element is only 32 parts per million, making its enrichment a difficult and expensive process, so that 1 mg of ^{46}Ca produced by electromagnetic separation may cost about £1000. This raises the question whether chemical methods of isotope separation might not prove cheaper than physical ones, and this question several institutions are attempting to answer.

As the result of the experiments in 1912–13 by F. A. Paneth in Vienna and G. de Hevesy in Manchester, both of whom demonstrated that Ra D (a lead isotope) and ordinary lead could not be separated by chemical methods, there was established a firm belief that isotopes show identical chemical behaviour and are therefore inseparable by chemical methods. In 1919 Lindemann (later Lord Cherwell) calculated that, in principle, isotopes must be separable, though the differences were so small that there would be little hope of success in practice. Throughout the 1920s many laboratories were studying methods of separation and achieved—mainly by distillation methods—minor changes in the isotopic ratios of elements such as neon, chlorine, zinc, and mercury. In 1932, however, H. C. Urey [3] showed that the electrolytic decomposition of water leads to a substantial enrichment of deuterium in the remaining fraction, and that repeated electrolysis could lead to essentially pure deuterium. Once it was realized that chemical methods could be utilized for isotopic separations, a large variety of methods was studied, and the fundamental techniques are now well understood. The quantities of isotopes now separated vary widely according to requirements; they range from grams per day for materials of interest for research—such as ^{10}B, ^{13}C, ^{15}N, ^{18}O—to kilograms per day for isotopes such as D and ^{235}U that are required for the atomic industries. ^{6}Li occupies an intermediate position.

Isotopic atoms have, of course, identical electronic configurations and therefore exhibit essentially the same chemical behaviour; they differ, however, in their atomic masses. In the case of the monatomic gases, this difference affects mainly the zero-point energies in the condensed phase, and the lighter isotope is therefore invariably more volatile. In the case of polyatomic molecules, however, vibration and rotation also contribute to the energy content, and may either reinforce or diminish the zero-point energy (that is, its internal energy at absolute zero); the combination in turn will affect volatility and thermodynamic activities. Thus, isotopic molecules, even if their molecular weight is the same—as in the case of HT and D_2—may differ appreciably in their vapour pressures at any given temperature (figure 1). The vibrational and rotational effects can more than counterbalance the mass effect; as has been shown by

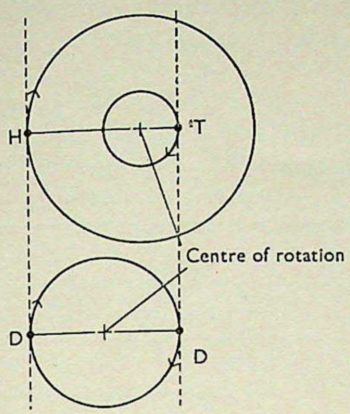


FIGURE 1 – *Effect of rotation of isotopic masses on molecular volume for HT and D_2.*

T. F. Johns, $^{12}C^{18}O$ is slightly more volatile than $^{13}C^{16}O$, even though the former is the heavier molecule [1*a*]. For example, at 60° K the ratio (α) of the vapour pressures of the two compounds is 1·0028; α is called the separation factor of the single-stage equilibrium. As isotopic mixtures are 'ideal solutions', the molar fractions (X) in the two phases of a vapour-liquid equilibrium are related by

$$\ln\left(\frac{X_1}{1-X_1}\cdot\frac{1-X_2}{X_2}\right)=\ln\alpha\simeq\alpha-1=\varepsilon$$

where ε is the separation parameter.

Isotope separation can in principle be achieved by two types of processes. These are respectively rate processes, in which one species reacts or diffuses faster than the other, usually as the result of a smaller mass, and equilibrium processes, in which one or the other species is thermodynamically favoured in one of the phases. To the first class belong separations by barrier diffusion, thermal diffusion, centrifugation, electrolysis, electro-migration, kinetic isotope effects, and, in a sense, even electromagnetic separation. To the second class belong distillation and chemical exchange reactions. When, as is usually the case, the difference of rate or equilibrium constants is only small, it is not possible to separate isotopes completely in a single operation; the operations may have to be repeated many times, in some cases thousands of times, before the desired degree of separation is achieved.

This need for repetition emphasizes one essential difference between rate processes and equilibrium processes. The former are irreversible reactions, and substantial amounts of energy have to be provided at every stage. The equilibrium processes, however, can be conducted in an almost reversible manner and, by making use of multi-stage counter-current processes, energy is required virtually only at the ends of the cycles. For comparable separations (that is with equal values of α), efficiently conducted equilibrium processes require only a very small fraction of the power required for rate processes.

It has been calculated that the minimum power-consumption W required to separate 1 mole of pure isotopic compound present originally at the molar fraction X is, for an irreversible process, $>4RT/\varepsilon^2X$. With a reversible process, with n theoretical plates in each column, this value, according to H. London [1*b*], need be only about $5{\cdot}3\ RT/n\varepsilon^2X$ (where $n \leqslant 500$), that is to say, two orders of magnitude less for the same conditions. But separation factors are not in fact equal for the two types of processes, and for very heavy isotopic molecules or atoms, for example UF_6, the separation parameter for irreversible processes (such as diffusion) greatly exceeds that for reversible processes such as distillation (Table 1). For relatively light atoms and molecules, however, reversible chemical processes show separation factors that are of the same magnitude as those arising from irreversible processes, and it is here that their practical applications become important.

TABLE 1

Isotopic species	Separation parameter ε	
	Distillation	Diffusion
H_2/HD	0·73	0·225
$^{10}BF_3/^{11}BF_3$	−0·0065	0·0074
$^{20}Ne/^{22}Ne$	0·0445	0·049
$^{16}O_2/^{16}O^{18}O$ . .	0·0052	0·030
$^{36}A/^{40}A$	0·006	0·055
$^{235}UF_6/^{238}UF_6$ . .	<0·0001	0·0043

The best-known example of a multi-stage reversible chemical process is fractional distillation, which involves the repeated equilibration between a gaseous and a liquid phase. In this case the chemical species in the gaseous and liquid phases are the same, for example,

$$\underset{\text{Liquid}}{^{20}Ne}+\underset{\text{Gas}}{^{22}Ne}\overset{\varepsilon=0{\cdot}044}{\rightleftharpoons}\underset{\text{Gas}}{^{20}Ne}+\underset{\text{Liquid}}{^{22}Ne}$$

though their proportions in the two phases vary, and refluxing is obtained simply by boiling at the bottom and condensing at the top of the column.

If the exchanging species are not of the same chemical type we have a gas-liquid exchange reaction, such as

$$\underset{\text{Gas}}{^{14}\mathrm{NO}} + \underset{\text{Soln.}}{\mathrm{H}^{15}\mathrm{NO}_3} \overset{\varepsilon=0{\cdot}055}{\rightleftharpoons} \underset{\text{Gas}}{^{15}\mathrm{NO}} + \underset{\text{Soln.}}{\mathrm{H}^{14}\mathrm{NO}_3}$$

To reflux, it becomes necessary to convert the NO to HNO_3 at the top of the column and the HNO_3 to NO at the bottom, as is described in more detail later. In general, chemical exchange reactions have considerably higher separation factors than distillation. For the distillation of nitrogen isotopes, $\varepsilon = P(^{14}N^{14}N)/P(^{14}N^{15}N) - 1 = 0{\cdot}0062$ at 64·8° K; this compares with $\varepsilon = 0{\cdot}055$ for the chemical exchange described above.

Other reversible processes may involve two immiscible liquid phases—as in solvent extraction; alternatively, one of the phases may be solid—as in ion-exchange and gas chromatography, the second phase then being liquid or gaseous.

DISTILLATION

The most important factor for the separation is the single-stage separation factor (α) or, rather, the parameter $\varepsilon = \alpha - 1$. The important column characteristics—like size and cost, power consumption, and equilibrium time—all depend on ε^2. In the case of vapour-pressure equilibria, α is the ratio of the vapour pressures of the two isotopic species. The value of ε depends on the temperature; below 100° K it is usually proportional to $1/T^2$. As the separation effect of a single-stage equilibration is small, a reflux column is used as counter-current cascade. For isotope separation the earlier type of column with bubble-caps is not very useful, because of the comparatively large height of the single stage and of its large retention of product. Columns filled with packing are nowadays employed, and this ensures continuous contact between the phases under more advantageous conditions. Such a column is considered to have *n* theoretical plates if its operation is equivalent to *n* successive equilibrations. The total change in isotopic concentration for a binary mixture, with molar fractions X and (1 – X) of the two isomeric species, between the top (T) and the bottom (B) of such a column at total reflux is then given by

$$\frac{X_T}{X_B} \cdot \frac{1-X_B}{1-X_T} = \alpha^n = \exp(\varepsilon n) \quad \text{if } \varepsilon \ll 1$$

The number of theoretical plates required to give a desired enrichment is shown in Table II.

TABLE II

Number of theoretical plates (n) required for various enrichment factors (R) and separation factors (α), assuming total reflux

R = overall enrichment factor	10	100	1000	10000
α	*n*			
2·0	3	7	10·0	13
1·4	7	14	21	27
1·1	24	48	73	97
1·05	47	94	141	188
1·02	115	230	345	460
1·01	230	460	690	920
1·005	460	920	1380	1840

When product is withdrawn—even at a low rate—the enrichment is considerably lower. This is demonstrated [1*c*] in figure 2, which illustrates the isotopic enrichment of $^{10}B(X_0 = 0{\cdot}2,\ \varepsilon = 0{\cdot}025)$. X is the molar fraction of ^{10}B in the product, and the abscissa gives the withdrawal rate P/B, where P is the withdrawal ratio of product and B is the boil-up rate in the column. The diagram shows that even with an infinitely long column the withdrawal rate P/B must be kept below 0·5 per cent, or $< 0{\cdot}2\varepsilon$, if a high purity of product is to be achieved.

The height of a theoretical plate depends greatly on the type of packing, which must simultaneously offer a large area of contact to the gas stream and the smallest possible retention of liquid so that equilibrium is quickly obtained. Theoretically, the smallest plate-heights should be achieved with very narrow empty tubes, of 1–2 mm internal diameter, in which the liquid runs down the wall in a thin film. This should give plate-heights equal to the tube diameter. But there are considerable practical difficulties in the use of such narrow wetted-wall columns; in particular the throughput is very low. This makes it necessary to use as many as 100 tubes in parallel if useful amounts of product are to be obtained [2]. For laboratory work it is more practical to employ wider tubes filled with a small random packing. By using cylindrically rolled wire netting—such as Dixon gauze packings (figure 4), which also give a low pressure-drop per plate—the height of a theoretical plate can be kept as low as 1 cm in

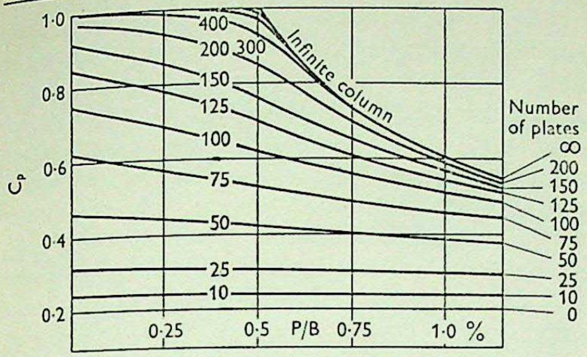


FIGURE 2 – *The effect of the withdrawal rate (P/B) on* ^{10}B *content of the product* (C_p) *during separation of boron isotopes* ($\varepsilon = 0{\cdot}025$).

small-diameter columns. For very large industrial columns, however, it is more economical to use expanded perforated metal sheet such as Spraypak (figure 5) or Panapak; though giving theoretical plate heights of about 45 cm, these allow much greater throughputs.

An interesting example of isotope separation by distillation is the production of heavy oxygen (^{17}O and ^{18}O) from water by I. Dostrovsky and A. Raviv [1*g*] at the Weizmann Institute of Science in Israel. Of these isotopes, the rare ^{17}O—which is contained in natural oxygen only to the extent of 0·037 per cent—is of particular interest, because its nuclear magnetic resonance is such as to permit non-destructive analysis, so opening up large fields of tracer application, particularly in biology and medicine. As the separation factor diminishes with increasing temperature, the distillation is carried out at reduced pressure and temperature. The plant operates in two stages. A pre-enrichment section, consisting of several four-inch columns 30 ft in length operating in parallel, produces an enriched product. This is then fed into a cascade of five columns, two of which are for stripping and three for further enrichment. This cascade represents more than 2000 theoretical plates, and the apparatus produces approximately 10 g per day of highly enriched $H_2{}^{18}O$.

EXCHANGE DISTILLATION

An interesting variation of the ordinary distillation process is the so-called exchange distillation, which can be employed if a liquid compound is dissociated in the gaseous phase, for example,

$$BF_3(C_2H_5)_2O \rightleftharpoons BF_3 + (C_2H_5)_2O$$

Here the factor which leads to the isotopic enrichment is not the difference between the vapour pressures of the compounds but the exchange reaction between gaseous BF_3 and the liquid $BF_3(C_2H_5)_2O$ complex:

$$^{10}BF_3(C_2H_5)_2O + {}^{11}BF_3 \rightleftharpoons {}^{11}BF_3(C_2H_5)_2O + {}^{10}BF_3$$

This combines the operational simplicity of a distillation process with the advantage of the considerably larger separation factor of the chemical exchange ($\varepsilon = 0{\cdot}034$ instead of 0·0075 as for distillation of BF_3). The method was first used in the United States in 1943, but details were not released until its rediscovery in Britain, Sweden, and Yugoslavia, independently, in 1957 [1]. The process is used technically for producing ^{10}B, for example, for neutron counters. A column operating in England contains about 400 theoretical plates and enriches ^{10}B, in kg quantities, from its natural content of 20 per cent to 96 per cent [1*c*].

In other examples of exchange distillation, the dissociation of trimethylamine acetate and of aqueous ammonium carbamate have been employed. The reactions are respectively:

$$CH_3.COONH(CH_3)_3 \rightleftharpoons N(CH_3)_3 + CH_3COOH$$
$$(\varepsilon = 0{\cdot}012 \text{ for } {}^{15}N \text{ enrichment})$$
$$NH_2.COONH_4(aq) \rightleftharpoons CO_2 + 2NH_3$$

In the second example both ^{15}N ($\varepsilon = 0{\cdot}015$–$0{\cdot}020$) and ^{14}C are enriched. In the case of the exchange distillation of BF_3-methyl-ether one finds in addition to the ^{10}B enrichment also a small enrichment of ^{18}O, arising from the exchange of the liquid compound with the gaseous ether ($\varepsilon = 0{\cdot}012$).

CHEMICAL EXCHANGE (GAS-LIQUID)

This method of isotope separation was first suggested in 1935 by H. C. Urey and L. J. Greiff, and led to the first 'large-scale' enrichment of ^{15}N, utilizing the exchange of gaseous NH_3 with solutions of ammonium salts ($\alpha = 1{\cdot}031$). The most elegant example of this type of process is, however, the exchange of HNO_3 with NO utilized by T. I. Taylor and W. Spindel in 1955 [1*d*], which also has a considerably larger separation factor. Using this process, L. Gowland and T. F. Johns [4] designed a cascade with two 15 ft columns in series, with a total of 300 theoretical plates, which enriched 0·4 g per day of ^{15}N from its natural concentration of 0·36 per cent to a purity of 99·8 per cent. An interesting point of their design was that it was fully automatic and did not involve the use of pumps or electronic devices. Each column had a refluxer at its base in which the nitric acid was converted to nitric oxide by means of sulphur dioxide, and a top refluxer where the nitric oxide was reconverted

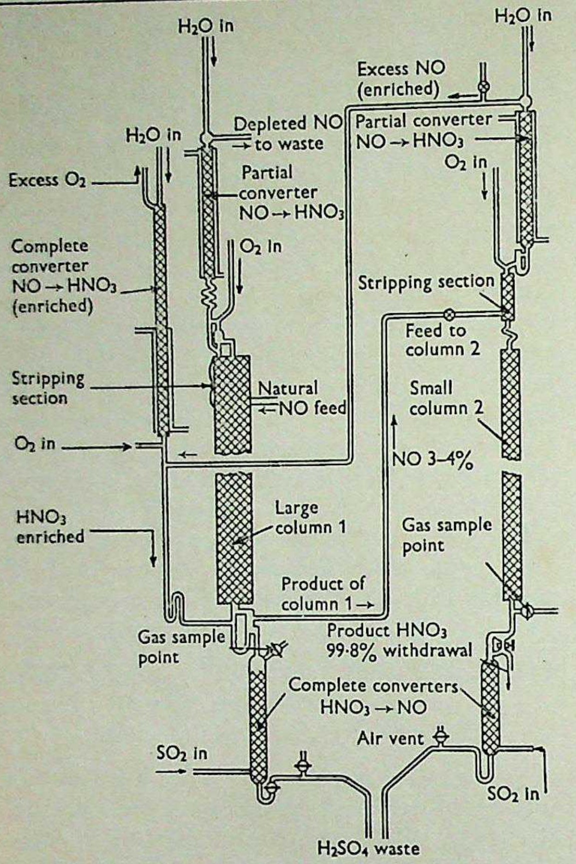

H₂O in
H₂O in
H₂O in
Excess NO (enriched)
Depleted NO → to waste
Partial converter NO → HNO₃
Excess O₂
Partial converter NO → HNO₃
O₂ in
O₂ in
Complete converter NO → HNO₃ (enriched)
Stripping section
Feed to column 2
Stripping section
Natural ⇌ NO feed
Small column 2
O₂ in
NO 3–4%
HNO₃ enriched
Large column 1
Gas sample point
Product of column 1 →
Gas sample point
Product HNO₃ 99·8% withdrawal
Complete converters HNO₃ → NO
Air vent
SO₂ in
SO₂ in
H₂SO₄ waste

FIGURE 3 – *Scheme of nitrogen isotope separation plant by Gowland and Johns. The arrangement is of particular interest because it uses two columns in series without the use of pumps.*

to nitric acid by combination with oxygen and water. The reflux ratio at the top of the columns could be adjusted simply by regulating the oxygen flow. As all transport between the columns took place via the gas phase, under pressure of the gases (figure 3), no pumps were required for either liquids or gases: this greatly increased the reliability of operation. The waste nitric oxide from the top of the second column was completely converted to nitric acid in an extra converter, and this still-enriched material was fed back to the first column.

Other chemical-exchange reactions with useful separation factors (at 25° C) are

$$S^{16}O_2 + 2H_2{}^{18}O \rightleftharpoons S^{18}O_2 + 2H_2{}^{16}O \quad (\alpha = 1{\cdot}020)$$
$$C^{16}O_2 + 2H_2{}^{18}O \rightleftharpoons C^{18}O_2 + 2H_2{}^{16}O \quad (\alpha = 1{\cdot}044)$$
$${}^{13}CO_2 + H^{12}CO_3{}^- \rightleftharpoons {}^{12}CO_2 + H^{13}CO_3{}^- \quad (\alpha = 1{\cdot}012)$$
$$H^{13}CN + {}^{12}CN^- \rightleftharpoons H^{12}CN + {}^{13}CN^- \quad (\alpha = 1{\cdot}012)$$
$${}^{34}SO_2 + H^{32}SO_3{}^- \rightleftharpoons {}^{32}SO_2 + H^{34}SO_3{}^- \quad (\alpha = 1{\cdot}019)$$
$${}^{36}SO_2 + H^{32}SO_3{}^- \rightleftharpoons {}^{32}SO_2 + H^{36}SO_3{}^- \quad (\alpha = 1{\cdot}040)$$

The design of chemical-exchange plants is very similar to that of distillation columns; Dixon rings and Helipak are usually the packing materials, giving the smallest theoretical-plate heights that can be attained. However, these heights never attain the small values obtained in the same columns for a distillation process. The reason for this is that the diffusion transport in the liquid phase—usually a solution of the exchanging molecule in water—is smaller by approximately an order of magnitude than in a pure liquid; also, the chemical exchange is often far from instantaneous.

CHEMICAL EXCHANGE BETWEEN TWO LIQUID PHASES

Isotope-separation processes involving equilibrium between two immiscible liquid phases are comparatively rare. Attempts in the author's laboratories to separate the lithium isotopes by exchange between a concentrated aqueous solution and a solution in tributyl phosphate did not show a significant separation factor. One case that has been described in the literature is the exchange of lithium between its liquid amalgam and various organic solutions. Thus G. M. Lewis and R. T. McDonald have investigated the amalgam exchange with solutions of lithium chloride in absolute ethyl alcohol and of lithium bromide in an alcohol-dioxan mixture. Reflux at the bottom of the column was produced by treating the lithium amalgam with alcoholic hydrochloric acid, thus transferring the lithium quantitatively into a neutral organic solution. Using an 18-metre glass column, they changed the $^7Li/^6Li$ ratio from 11·6 at the top (feed-point) to 5·5; their data do not permit a calculation of the single-stage separation factor [5]. However, a more recent report by L. Perret, L. Rozaud, and E. Saito [6] gives the single-stage separation factor between Li-amalgam and a solution of lithium bromide in dimethylformamide as approximately 1·05, and the values of α for other organic solvents may be expected to be very similar.

DUAL-TEMPERATURE CHEMICAL EXCHANGE

In the processes described so far—using distillation or exchange—reflux could be obtained either by condensation or by simple chemical processes. But in a number of exchange reactions, which are otherwise very useful for isotopic enrichment, no such simple refluxing process is possible. A typical case is

$$HOH + HSD \rightleftharpoons HOD + HSH$$

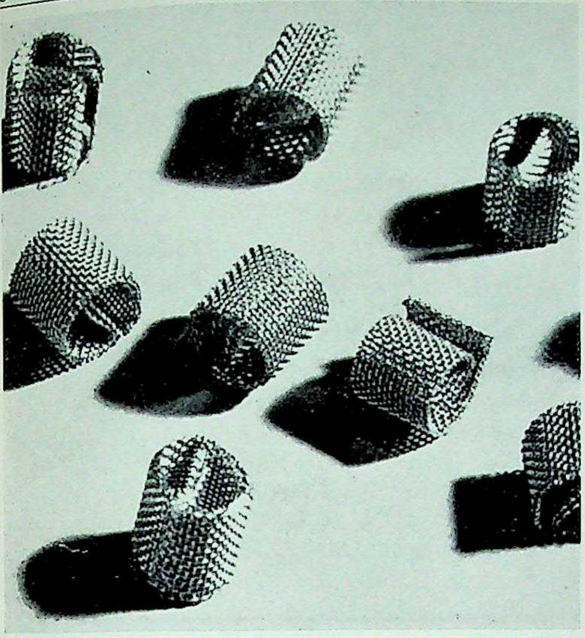


FIGURE 4 – *Dixon or Lessing gauze rings are among the most efficient packing materials for laboratory or pilot plant columns.*

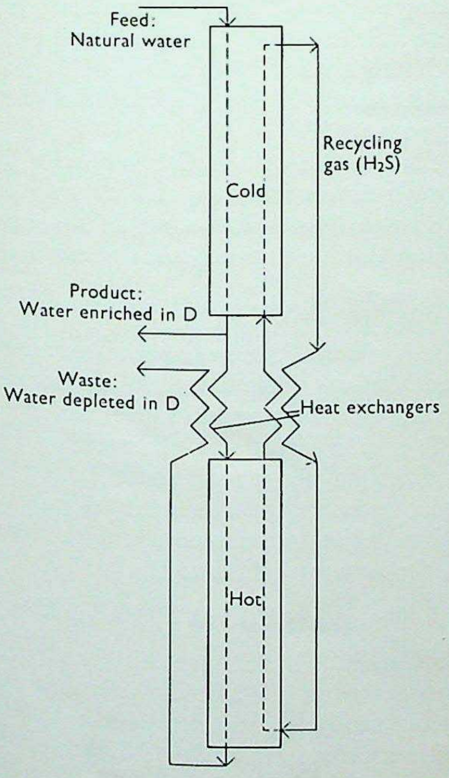


FIGURE 6 – *Schematic representation of dual-temperature chemical exchange process, for cases where conversion of one chemical species into the other at the refluxing ends of the column is not practicable.*

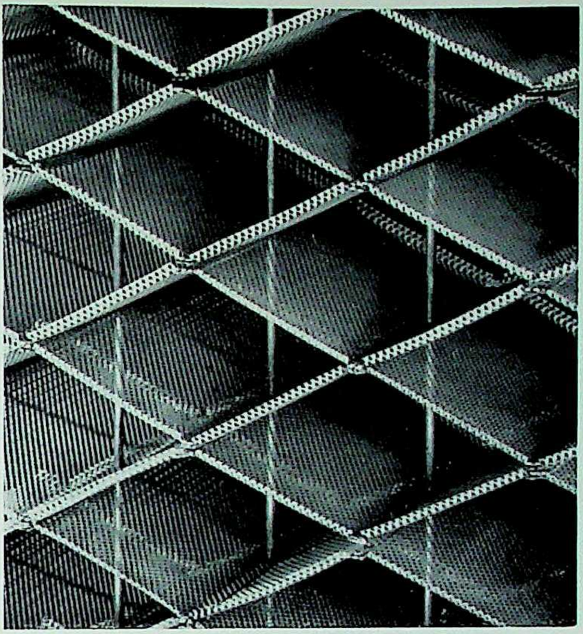


FIGURE 5 – *Spraypak is one of the most efficient packing materials for large-scale columns for bringing liquids and gases into intimate contact.*

which requires at both ends of the column the complete conversion of hydrogen from H_2S to H_2O and *vice versa*.

In these cases, the refluxing action can be provided by a second column, running at a higher temperature (with a correspondingly lower value of α) which provides the enriched H_2S; this is fed against the current of H_2O into the enrichment end of the 'cold' column (figure 6). A simplified scheme can be visualized in which water of natural D_2O content (0·0145 per cent) is fed into the top of the cold column, which at 30° C has an effective separation factor $\alpha = 2{\cdot}27$; the waste, with reduced D_2O content, is withdrawn from the bottom of the hot column, which, at 130° C for example, has an effective separation factor $\alpha = 1{\cdot}65$. The actual equilibrium constants are somewhat different, because the separation factor is modified by the presence of steam in the gas phase, largely H_2S, and by the solubility of H_2S in water. Enriched product is withdrawn, as water, between the two columns. In practice, several dual-temperature systems are coupled in series, the depleted waste from the second system being returned to the first one. Generally, a product containing 10–20 per cent of D_2O is produced by this method, and this is finally concentrated by other methods, for example electrolysis or distillation.

This process, patented by J. S. Spevack [7] though proposed independently by Rideal and by

Harteck and Suess, was primarily designed for production of deuterium on the largest scale, for use in heavy-water reactors. It is very economical in power consumption, the demand arising mainly from the pumping power required to circulate the liquid and gas and the heat required to humidify the gas stream when entering the hot column. The energy required for heating the gas and liquid before they go into the hot column is largely recovered by heat exchange with these fluids when they leave the hot column. The temperature of the cold tower must be above 29° C, as otherwise H_2S forms a solid hydrate; that of the hot tower should be below 130° C, as otherwise the proportion of steam in the H_2S phase would become too high, which would adversely affect the separation process and the power consumption.

Similar dual-temperature processes have been proposed using the exchange of liquid H_2S with liquid H_2O; for the exchange of gaseous hydrogen with liquid NH_3, with potassamide as a catalyst [1*e*]; and for the exchange of gaseous hydrogen with water, using suspensions of platinum and carbon as exchange catalysts.

Though there would appear to be a wide field of application for dual-temperature processes, there are in fact severe practical limitations. The ratio of the mass-flows in the gaseous and liquid cycles must lie between the values of α at the two temperatures, and for all but the largest separation factors this entails a difficult control of the flow ratio.

CHROMATOGRAPHIC SEPARATIONS

While counter-current exchanges with two fluid phases (gas-liquid or liquid-liquid) have found application on the largest scale, the use of chromatographic processes, that is of exchangers involving one solid and one fluid (gas or liquid) phase, has not led to large-scale technical applications in isotope separation, although several interesting scientific investigations have resulted from them.

A priori, one would have thought that chromatographic methods would be very useful for isotope separations, because of the very small plate-heights (<1 mm) that can be achieved in such columns. This would make it possible, without much expense, to operate with 10 000 and more theoretical plates in a single column as compared with the 500-odd theoretical plates that the best distillation column can provide. This advantage however, is more than offset by the extreme slowness with which equilibria that involve exchange between a solid and a liquid phase are attained, and by the very small number of gas-solid exchanges that give a reasonable separation factor.

The reactions involving a liquid phase can be subdivided into two groups:

1. Straightforward ion-exchange chromatography between an electrolyte solution and a solid ion-exchanger. Here the separation factors observed are all exceedingly small. Even for the most favourable case of $^6Li/^7Li$ the observed values of ε lie between 0·0010 and 0·0038, rising with the degree of cross-linking of the exchanger [8]. When we come to heavier ions, such as sodium, values around 0·0001 have been observed [9] for the exchange of ^{22}Na and ^{24}Na.
2. Exchange between the ions in the exchanger and an associated compound in the aqueous phase. The best-known example of this type is the enrichment of ^{15}N obtained by F. H. Spedding, J. E. Powell, and H. J. Svec [10], who used the equilibrium

$$NH_4^+ \text{ (exchanger)} \overset{\varepsilon=0{\cdot}026}{\rightleftharpoons} NH_4OH(aq)$$

After recycling a band of NH_4^+ through a series of 10 ft columns, so that the band travelled through a total length of more than 300 ft, they were able to produce, from a starting material with the natural ratio of $^{15}N/^{14}N = 0{\cdot}0036$, a small amount of enriched material containing 45 per cent ^{15}N.

The exchange reaction effective for the isotope separation is in this case not the ion exchange (which has a negligible separation factor) but the aqueous-solution equilibrium

$$[NH_4^+](aq) + [OH]^- \rightleftharpoons NH_4OH$$

The function of the exchanger is merely to hold back the ions while the solution containing the undissociated ammonia moves on.

A similar effect was observed by Japanese workers [11] for the passage of boric acid over anion exchanger. Here the effective equilibrium is the first dissociation of the boric acid in the aqueous solution:

$$B(OH)_3 + H_2O \rightleftharpoons B(OH)_4^- + H^+$$

For the $^{10}B/^{11}B$ exchange this gives a value of ε between 0·010 and 0·016. Again the exchanger serves merely to separate the ionic and non-ionic species from each other and so to produce the multi-stage effect.

In the case of gas chromatography, separation factors differ considerably according to the nature of the adsorption process involved. Utilizing physical adsorption, only very small values of ε have been observed: for example, for the separation of the neon isotopes with charcoal columns at liquid-nitrogen temperature ($\varepsilon = 0{\cdot}0020$) [12]. However, the untried case of $^3He/^4He$ might well give good separations, if adsorption on a charcoal column at liquid-hydrogen temperature is employed, as a result of the large mass-ratio of the isotopes and of the low temperature of separation.

If chemical adsorption is utilized, as in the gas-chromatographic separation of the hydrogen isotopes on palladium columns by E. Glueckauf and G. P. Kitt [1*f*], very large separation factors arise: for example, $\alpha = 1{\cdot}8$ for H—D and $\alpha = 2{\cdot}6$ for H—T mixture. The attainment of local equilibrium is fast, and, consequently, quick and efficient separations can be carried out; enrichment by a factor of 100 is easily obtained for substantial quantities of material.

If one is interested only in the heavier hydrogen isotope, it is enough to fill about a third to a half of the column with the isotopic mixture and then to push this band through the column either with further quantities of ordinary hydrogen or by progressively heating the column from the rear. As deuterium or tritium are less readily adsorbed on palladium than is hydrogen, they form a frontal band of high purity that can easily be separated from the rest. The sharpness of the separated bands can be seen from figure 7.

Unfortunately, the method is not economical for large-scale production of deuterium, where not simplicity but power consumption becomes the important factor. This is partly because of the high heat of adsorption of hydrogen on palladium, heat that has to be supplied for the regeneration of the column. A contributory factor is the use of gaseous hydrogen as the starting material; this is rather more expensive than the water used for the H_2S process described above. But for the reconcentration of litre quantities of dilute deuterium or tritium mixtures the method is unchallenged, and very much cheaper than, for example, thermal diffusion methods.

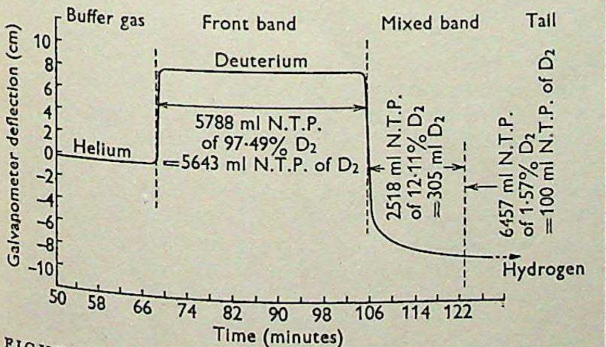


FIGURE 7 – *The chromatographic separation of hydrogen isotopes on a column containing 460 g of palladium powder. The graph gives the catharometer deflections of the eluted gas: first the helium buffer gas originally in the column, then the deuterium fraction, followed by the hydrogen fraction.*

CONCLUSION

As the chemical properties of different elements vary widely, so also do the chemical methods appropriate for separating them into their isotopes. Table III gives a short survey of the most promising methods that can be used for the lighter isotopes, excluding hydrogen; most of these have been tested either by laboratory methods or on a considerable scale. In the case of the hydrogen isotopes, the separation factors are so large that a variety of methods are possible; the choice will then depend on the quantities required, the available starting product (H_2, NH_3, or H_2O), and a number of economic factors.

Above an atomic weight of 40, chemical isotope

TABLE III

Element	Most promising method of isotope separation	α
$^3He/^4He$	Small-scale gas chromatography on charcoal cooled with liquid H_2	Unknown
$^6Li/^7Li$	Liquid-liquid chemical exchange between Li-amalgam and, for example, solution of dimethyl formamide in $LiBr_2$	1·05
$^{10}B/^{11}B$	Exchange distillation $BF_3 \rightleftharpoons BF_3(C_2H_5)_2O$	1·027
$^{13}C/^{12}C$	Distillation of CO	1·011
	Chemical exchange between HCN and CN^- (gas-liquid) ..	1·012
$^{15}N/^{14}N$	Chemical exchange (gas-liquid) $NO \rightleftharpoons HNO_3$	1·055
$^{18}O/^{16}O$	Distillation of NO .. .. or H_2O .. .. or CO	1·046 1·006 1·0076
$^{22}Ne/^{20}Ne$	Distillation of Ne	1·044
$^{34}S/^{32}S$ $^{36}S/^{32}S$	Chemical exchange (gas-liquid) between SO_2 and HSO_3^- ..	1·019 1·040
$^{36}A/^{40}A$	Distillation	1·006

separation is—at least at present—inferior to physical methods such as electromagnetic separation and diffusion. This inferiority is due both to the smallness of the chemical separation factors, which fall off rapidly with increasing atomic number and to the lack of pairs of compounds in which the chemical bond is sufficiently different and which also rapidly establish equilibrium with each other; this lack is particularly felt in the case of the metallic elements.

REFERENCES

[1a] JOHNS, T. F. 'Proceedings of the International Symposium on Isotope Separation'. North Holland Publishing Co., Amsterdam, p. 174. 1958.

[1b] LONDON, H. *Ibid.*, p. 319.

[1c] MCILROY, R. W. and PUMMERY, F. C. W. *Ibid.*, p. 178.

[1d] TAYLOR, T. I. and SPINDEL, W. *Ibid.*, p. 158.

[1e] BIGELEISEN, J. *Ibid.*, p. 121.

[1f] GLUECKAUF, E. and KITT, G. P. *Ibid.*, p. 210.

[1g] DOSTROVSKY, I. and RAVIV, A. *Ibid.*, p. 336.

[2] KUHN, W., BAERTSCHI, P., and THURKAUF, M. *Chimia*, 8, 109, 145, 1954.

[3] UREY, H. C. and WASHBURN, E. W. *Proc. Nat. Acad. Sci., Wash.*, 18, 496, 1932.

[4] GOWLAND, L. and JOHNS, T. F. AERE C/R 2629, Harwell. 1958.

[5] LEWIS, G. M. and MCDONALD, R. T. *J. Amer. chem. Soc.*, 58, 2519, 1936.

[6] PERRET, L., ROZAUD, L., and SAITO, E. Commissariat à l'Energie Atomique, Rep. No. 1029, 1958.

[7] SPEVACK, J. S. Columbia University Rep. A-393, 1942. Also MDDC-891, 1947.

[8] LEE, D. A. and BEGUN, G. M. *J. Amer. chem. Soc.*, 81, 2332, 1959.

[9] BETTS, R. H., HARRIS, W. E., and STEVENSON, M. D. *Canad. J. Chem.*, 34, 65, 1956.

[10] SPEDDING, F. H., POWELL, J. E., and SVEC, H. J. *J. Amer. chem. Soc.*, 77, 6125, 1955.

[11] YONEDA, Y., USHIJIMA, T., and MAKISHIMA, S. *J. phys. Chem.*, 63, 2057, 1959.

[12] GLUECKAUF, E., BARKER, R. H., and KITT, G. P. *Faraday Soc. Disc.*, 7, 199, 1949.

Book reviews

PHYSICS

A Physics Anthology, *edited by N. Clarke. Pp. viii+323. Published on behalf of the Institute of Physics by Chapman and Hall Ltd, London; Reinhold Publishing Corp., New York.* 1960. *35s. net.*

This book consists of twenty-two articles published in the 'Bulletin' of the Institute of Physics and mostly based on lectures and addresses delivered at meetings organized by the Institute. It concludes with a report of a discussion dealing with 'Ethics and the Scientist' that was held in Manchester in 1956. Considerations of space will not allow even a list of the distinguished men responsible for the contributions nor of their individual subjects, which cover a wide field. A number of the contributions deal with physics as a factor in education and with the education of physicists; others deal with different aspects of physics in relation to industry; others again with the history of the subject, in particular two articles on Newton and Max Planck. They are all written by experts, some of them world-famous, with excellent powers of exposition. The collection should have a wide appeal.

It is worthy of special note that the book opens with an address by Sir Cyril Hinshelwood on 'Physics among the Sciences', written with characteristic conciseness, charm, and insight. This contains the sentence: 'What any man may believe about the ultimate purpose of existence can be little affected by his status in the world of space time', leading to the point that relativity has nothing to do with religion. Canon Raven, however, in an article on 'The Impact of Physics on Science and Religion', seems to imply that Einstein's theory of relativity had a profound influence on the religious and philosophic outlook. Perhaps Einstein may be allowed a word on the subject. It is authoritatively recorded that the Archbishop of Canterbury asked him whether he thought that his theory had any religious application and that the reply was a blunt 'None'. This is an example of how great men of science disclaim for their work the implications that theologians find in it.

E. N. DA C. ANDRADE

CONDUCTION IN SOLIDS

Electrons and Phonons; The Theory of Transport Phenomena in Solids, *by J. M. Ziman. Pp. xiv+554. Clarendon Press, Oxford; Oxford University Press, London.* 1960. *84s. net.*

When electricity travels through a solid it is transferred, in most cases, by electrons that move through the solid. Heat is in the main transferred by small-scale mechanical vibrations; in solids that conduct electricity, heat is also carried by the electrons. These phenomena require the quantum theory for their proper understanding, and one of its main consequences is that vibrational energy occurs in quanta. These behave in such a way that it is convenient to apply the language of particles to them, and to give them a name of their own, phonons. The theory of electrical and thermal conduction is thus largely concerned with the behaviour of electrons and phonons. These have many common features, and this fact makes a unified treatment illuminating.

In this book Dr Ziman sets out to explain the theory of transport phenomena in solids in a way that stresses the physical principles. He makes no pretence that the theory is everywhere on a rigorous footing. Rather, he exposes every dubious assumption for what it is, and leaves one to appreciate more clearly the foundations on which such an elaborate edifice has been built. He covers all the major aspects of conduction in solids, and deserves our gratitude for collecting in one place so much that would otherwise be difficult to find. The book is very readable; whenever the argument is going to run into turgid mathematics, Dr Ziman is at pains to give a sketch-map of the forest of his argument before we get lost in what he calls the undergrowth of suffixes. M. H. L. PRYCE

INORGANIC CHEMISTRY

Modern Aspects of Inorganic Chemistry (*third edition*), *by H. J. Emeléus and J. S. Anderson. Pp. xi+611. Routledge and Kegan Paul Ltd, London.* 1960. *35s. net.*

This book needs no introduction to chemists, most of whom are already greatly indebted to the authors for an important part of their chemical education. Interest in this revised edition—which has largely been rewritten and reset—naturally centres around the changes that the authors have made. Their task of selection is not an easy one, as there has been so much rapid development in inorganic chemistry during the past decade.

There has been only a modest increase (about 10 per cent) in size, and the authors are to be complimented on their choice of material for deletion and addition. Under the heading 'π-Bonded Complexes' there is an up to date discussion of carbonyls, nitrosyls, cyclopentadiene complexes, and metal-aromatic complexes generally. The rapidly developing subject of inorganic-reaction mechanisms receives full attention. The discussion of modern valency-theory has been extended to include ligand field-theory and a little more on the molecular-orbital approach. The discussion of bonding in metal complexes embraces most recent developments.

This book should be in the hands of every teacher of chemistry and will do much to encourage the modern physical approach to the subject. So far as its value to the undergraduate is concerned it should be prescribed reading—I cannot give higher praise than to say that it is as good as before, but more up to date.

The authors have done an excellent job, and it is fitting to compliment the publishers on the fact that the price is still the same as that of the 1952 edition.

R. S. NYHOLM

ORGANIC CHEMISTRY

Advances in Organic Chemistry: Methods and Results, *Vol. I, edited by R. A. Raphael, E. C. Taylor, and H. Wynberg. Pp. ix+387. Interscience Publishers Inc., New York; Interscience Publishers Ltd, London.* 1960. *90s. net.*

This volume is the first of a new series which, according to the general introduction, will contain 'articles providing a critical appraisal and evaluation of new aspects ripe for further development and of novel extensions to well established methods'. The contents of volume I are: the Kolbe electrolytic synthesis (B. C. L. Weedon); polyphosphoric acid as a reagent in organic chemistry (F. Uhlig and H. R. Snyder); the Wittig reaction (S. Trippett); hydroxylation methods (F. D. Gunstone); the selective degradation of proteins (E. O. P. Thompson); and optical rotatory dispersion and the study of organic structures (W. Klyne).

Although most of the topics in this volume have already been reviewed at least once in the last five years, the scope and treatment of the first four topics are different from those of any previous reviews and are comparable with those that appear in *Angewandte Chemie*. They include examples of typical experimental procedures. The last two articles, occupying more than half the book, describe the very great progress made in these subjects during the last few years. The new series is welcome, and can be recommended to all organic chemists. J. F. W. McOMIE

SALT MARSHES

Salt Marshes and Salt Deserts of the World, *by V. J. Chapman. Pp. xvi+392. Leonard Hill (Books) Ltd, London; Interscience Publishers Inc., New York.* 1960. *95s. net.*

Professor Chapman has successfully integrated many important studies on the physiography and vegetation of salt marshes and alkali deserts. Saline habitats, with the exception of mangrove swamps, are treated on a world scale.

The author's survey falls into three main parts: the first deals with the characteristics, formation, and ecological factors of salt marshes and inland salt deserts; the second with their vegetation; and the third with the physiology and biology of halophytes. Successional relations of the vegetation are summarized in many useful schemes, possibly somewhat dogmatically in view of the conjectural nature of the relationships in at least some of the areas. There are interesting sections on life-form analysis and on the biogeographical relationships of salt-marsh plants. The algal communities are given special attention.

This monograph should serve both as a source of information, and as a stimulant, to those studying salt-marsh vegetation. The book is abundantly illustrated and adequately documented, although more explicit legends to some of the figures and tables would help the reader. The photographs are a pleasing feature. A. J. WILLIS

GEOLOGY

Lead Isotopes in Geology, *by R. D. Russell and R. M. Farquhar. Pp. viii+243. Interscience Publishers Inc., New York; Interscience Publishers Ltd, London.* 1960. *65s. net.*

It is now well known that it is possible to estimate the age of the earth's crust with some approach to precision by using the isotopic constitution of lead from galena and other sources, and that in favourable cases the approximate ages of lead-ore deposits can be determined. Some guidance is also afforded in unravelling the more complex ancestry of ores containing anomalous leads, that is those with an excess or deficiency of radiogenic lead relative to their ages. The authors of this most useful monograph are to be congratulated on their enterprise and success in bringing together in an attractive and easily accessible form all the information now available—hitherto widely scattered and in part previously unpublished—on a subject that has far-reaching applications in geology and geochemistry. After a general introduction, the successive chapters deal with methods, age of the earth, dating galenas, anomalous leads, case histories, extensions of the Holmes-Houtermans model, and lead-uranium-thorium methods. Tables of data, constants, and functions required for various calculations are given. The authors favour the hypothesis that the source of lead for lead ores is the earth's mantle, and that the lead has been carried to the surface, probably by volcanoes during the processes of continent formation; but they recognize that there are objections to this view and the model based on it, and in practice they deal quite fairly with all the models so far proposed. No one has yet given sufficient attention to the mixing and differentiating consequences of erosion, geosynclinal sedimentation, and subsequent granitization. A particularly useful feature is the appendix, which comprises several hundred isotopic analyses of common leads.

ARTHUR HOLMES

CYBERNETICS

Design for a Brain (*second edition, revised*), *by W. Ross Ashby. Pp.* ix + 286. *Chapman and Hall Ltd, London.* 1960. 42*s. net.*

Dr Ashby has added as a subtitle to the new edition of his book, 'The origin of adaptive behaviour'. 'Origin' may seem to the biologist rather ambitious, as the author deals hardly at all with the variety of organisms. What he does is to discuss the logic of 'trial and error machines', a subject that is obviously of the greatest biological interest. The novelty of his contribution is that he tries to show how trial and error, working through a series of 'step functions', makes possible even the highest activities of the nervous system. Indeed, he believes that trial and error is the necessary process that alone is capable of eliciting from the environment the information that is required for homeostasis. He is not disposed to look very closely for the actual origin of the variety that he postulates. He hints that the various steps may be found among unspecified events, at the molecular level, in neurons.

Not all biologists will readily follow the logic of Dr Ashby's proofs, but perhaps those least inclined to do so could learn most from attempting it. Biological problems involve many concepts that are difficult to make precise: one such is adaptation. We should be grateful to those who are prepared to do the hard thinking that is needed. The illustration of Dr Ashby's concepts is drawn, in the main, not from actual living organisms but from the model he has built, the homeostat. It is disappointing that the curves he provides from the working of this machine are the same in the second edition as the first. He does not seem to have been able to draw further information from it in the intervening seven years. We need an analogue that will compute continually for us parameters that we need to know—in fact, one that works as the brain does. In the meantime, Dr Ashby's analysis may encourage us to look at neural mechanisms with less rigid preconceptions as to what the significant variables may be.

J. Z. YOUNG

REPRODUCTION AND FERTILITY

Journal of Reproduction and Fertility, *Vol.* 1, *No.* 1, *February* 1960, *edited by C. K. Austin. Pp.* 120. *Blackwell Scientific Publications, Oxford. Quarterly,* 20*s. net; annual subscription,* 70*s. net.*

The promoters of the 'Journal of Reproduction and Fertility' are to be congratulated on the first number, which contains many interesting papers. There is an important article on phospholipids in mammalian semen, by Hartree and Mann. This subject has been confused for some time so far as mammalian spermatozoa are concerned: Hartree and Mann's clarification of the situation is therefore of special interest. Beatty's paper on the fertilizing capacity of semen obtained from different rabbits is also important; it is particularly welcome as a continuation and clarification of Russian work on heterospermic insemination. Bruce's paper on the effect of strange male mice in inhibiting pregnancy is of unusual interest. Clinicians and research workers in reproduction and gametology will value the possibility they now have of reading papers in their own fields in one journal, instead of in ten or more.

I hope the editors will publish papers on reproduction and gametology in plants and in animals other than mammals, because much was learned and is still to be learned from such studies. There is a tendency for the mammalian gametologist to pay too little attention to his invertebrate and botanical counterparts, and *vice versa.*

ROTHSCHILD

GENETICS

Human Biochemical Genetics, *by H. Harris. Pp.* viii + 310. *Cambridge University Press, London.* 1959. 37*s.* 6*d. net.*

In his book Dr Harris presents an admirable account of the remarkable advances in the study of human biochemical genetics since the publication of the second edition of Garrod's 'Inborn errors of metabolism' in 1923. He has condensed in an orderly and coherent manner information that is scattered throughout the literature. He starts off with two useful chapters in which, besides emphasizing the value of Garrod's concept of the biochemical defect, he recalls the basic ideas of genetics and provides a mathematical background to population genetics. Subsequently he discusses, mainly from a biochemical viewpoint, the well recognized and newly discovered normal and abnormal variations of the metabolism of amino-acids, carbohydrates, and other substances, and of the constitution of blood. The two chapters on human haemoglobins and blood-group substances are outstanding for their clarity. Except on appropriate occasions he avoids speculative or provocative parts of the subject, and also avoids mentioning species other than man, although he carefully indicates the weaknesses of accepted views and the problems yet to be answered. With this mode of approach, and with an ability to write clearly and concisely, he has produced a readable book, which is as up to date in its information as the speed of publication will allow.

J. R. P. O'BRIEN

ECOLOGY

The Biology of Weeds, *edited by J. L. Harper. Pp. xv* + 256. *Blackwell Scientific Publications, Oxford.* 1960. 42*s. net.*

For many years there has been a scarcity of books dealing with the biological side of weed studies as compared to the mass of literature dealing with their control; possibly weeds were not considered respectable for academic work. This volume will help to correct the position: it is a record of some of the papers presented at a symposium organized by the British Ecological Society at Oxford in 1959, a meeting convened to bring together research workers in the fields of taxonomy, evolution, genetics, physiology, and ecology of weeds. Although of special interest to the ecologist, this book will undoubtedly be of value to the applied biologist, and should prove useful to the worker in weed control.

The topics covered are diverse, and include well documented contributions on the history of weeds in Britain; dormancy in weed seeds; the spread of weeds by crop seeds; plant competition; the ecology of weeds; toxin production by couch-grass; the comparative life histories of creeping thistle and coltsfoot; and the germination and early establishment of the plantains. Twenty-four papers are given, and Dr Harper's introduction to the volume forms a concise summary of the aim and scope of the symposium.

D. S. C. ERSKINE

PLANT PATHOLOGY

Plant Pathology, *Vol. II,* The Pathogen, *edited by J. G. Horsfall and A. E. Dimond. Pp. xiv* + 715. *Academic Press Inc., New York.* 1960. $22.

This volume, the second of three in the series, discusses the antecedents of disease, how disease originates, and, in short, what causes disease. There are chapters on, among other subjects, the pathogen; the nature, origin, and evolution of parasitism; the multiplication and inactivation of viruses; the reproduction of bacteria, actinomycetes, and fungi; spore germination; the breaching of host barriers by mechanical and chemical means; fungitoxicity and fungicidal chemistry; and nematocides.

Each author is a specialist in his own field, and the style is very readable throughout. It is impossible for a present-day plant pathologist to be familiar with advances in all the aspects of his subject, but this book will give him a comprehensive summary of some. C. C. V. BATTS

PHARMACOLOGY

Introduction to Pharmacology, *by J. J. Lewis. Pp. xii+826. E. & S. Livingstone Ltd, Edinburgh.* 1960. 55*s. net.*

Pharmacology is a troublesome subject because it can be approached from two quite different directions, from that of chemistry and that of biology. The author of this book has approached it from the chemical side, and for the chemist this is an admirable book. For the biologist, and therefore for the medical student, it is, understandably, less suitable. The section on the fate of drugs after absorption, dealing with the chemical changes they undergo, is interesting, informative, and concise; but, on the other hand, the section on the sympathomimetic amines and their physiological properties is out of date.

The chemical emphasis is constantly seen throughout the book. In describing the corticosteroids the author has been fascinated by the new compounds that have been recently introduced and that have shown steadily increasing potency. He has, however, nothing to say about the function of aldosterone, except to mention it in one figure as a very potent minerocorticoid.

The book is intended for medical students; it is to be feared that it will appeal strongly only to the occasional one who has a flair for chemistry.

J. H. BURN

PHARMACY

Bentley and Driver's Textbook of Pharmaceutical Chemistry (*seventh edition*), *revised by J. E. Driver. Pp. ix+728. Oxford University Press, London.* 1960. 63*s. net.*

During the thirty-five years that have passed since the first edition of this book appeared, professional pharmacy has entirely changed, from a craft, modified by introduction of the basic sciences, to a full academic discipline based on the application of the basic sciences to modern pharmaceutical techniques. The book attempts the impossible: the boundaries of relevant knowledge have been so far extended that it would need many volumes to cover the field.

The new edition is vastly superior to its predecessors, and it is only fair to say that the impossible task has been attempted with some success. Part I covers analytical methods, but many chapters are too short to be of use. Part II, on inorganic chemistry, introduces two new chapters on atomic nuclei and radioactivity, and on periodicity and valency, that are as good as they can be in relation to the space allowed, but would have been much improved had it been possible to include the basic knowledge that is necessary to predict the types of bonds present in inorganic compounds. It is a pity that the study of the modern aspects of inorganic chemistry has not been taken further.

Two-thirds of the book (Part III) is devoted to organic chemistry, and has been extensively revised. Too little attention has been paid to the fundamental roles of inductive, mesomeric, and tautomeric effects, with the result that the student may well be confused.

The factual matter concerning medicinal organic compounds is very well done and mainly up to date; it would be carping to mention certain details with which one did not agree. On the other hand, it is a matter of regret that the fundamentals such as ionization and factors influencing drug action are not discussed at all. Nevertheless, the book would form a valuable addition to the library of anyone interested in pharmaceutical chemistry.

W. H. LINNELL

HISTORY OF SCIENCE

A History of Greek Fire and Gunpowder, *by J. R. Partington. Pp. xvi+381. W. Heffer & Sons Ltd, Cambridge.* 1960. 70*s. net.*

Professor Partington has made a magnificent contribution to science and to history by bringing his vast knowledge of chemistry to bear on the story of firearms, in the widest sense of the word, from their first appearance into modern times. The range of his inquiry stretches from Western Europe to China, and he has studied all the available sources. Only trained chemists are fitted to dispute his conclusions. Lay historians will be particularly interested by his convincing analysis of Greek Fire and of the part played in it by petroleum, and by his estimate of the role of the Arabs in the development of gunpowder. Apart from some unimportant errors over dates and names and some questionable historical statements, his scholarship is as meticulous as it is wide. One or two of the translations that he uses could be criticized, and he is over-scrupulous in mentioning every possible commentary on his original sources. The subject also raises economic problems that need fuller discussion. But his book will long remain the essential and definitive compendium of information about the use of chemical substances in warfare in past ages. S. RUNCIMAN

SCIENCE IN WARTIME

The Sciences were never at War, *by Sir Gavin de Beer. Pp. xv+279. Thomas Nelson & Sons Ltd, London.* 1960. 30*s. net.*

Edward Jenner wrote to the *Institut National* of France, in about 1803, a letter containing the sentence 'The Sciences are never at war'; and Sir Gavin de Beer, who is equally at home in English and French culture and letters, has adapted these words as the title for his book. This is concerned with the relationship between English and French men of science from the early days of the Royal Society to the fall of Napoleon. During this period the two countries were at war for an aggregate of nearly sixty years; yet, as the book clearly demonstrates, there was always free and cordial communication between the men of science, culminating in the celebrated visit of Sir Humphry Davy to France, by special permission of Napoleon, in 1813, when he was most hospitably entertained.

The book consists mainly of letters written by celebrated scientific figures, such as Réaumur, Sir Joseph Banks, Delambre, and Jenner, as well as by men of less note, concerning the scientific relationship of the two countries during periods of war, with sufficient felicitous comment to give the reader the contemporary background. Of particular interest is the period of the Napoleonic wars, when the efforts of Sir Joseph Banks played so considerable a part in maintaining friendly contact between the scientists of the two countries. Most of the letters and documents have never been printed before. Translations are provided of all the letters written in French. The book is a valuable contribution to an aspect of the history of science that has not been stressed before. E. N. DA C. ANDRADE

HENRY CAVENDISH

Henry Cavendish: His Life and Scientific Work, *by A. J. Berry. Pp.* 208. *Hutchinson, London.* 1960. 35*s. net.*

This may perhaps be considered Mr Berry's best book; he has performed a useful service by portraying the full extent of Cavendish's amazingly precise scientific work, much of which was not published during his lifetime. The 'Life', however, is little more than a post-mortem examination, as is implied by the heading 'Ancestry and personal characteristics'. A sketch of contemporary developments is then given as an introduction to chapters on Cavendish's researches on general chemistry, dynamical subjects, electricity, heat, meteorology, and the composition of water.

Though useful and instructive for the general reader and the working scientist, the book cannot be unreservedly recommended to students of the history of science. In one place the author betrays his attitude by dismissing a topic as being 'only of historical importance'. What are we to conclude about the remainder? With Cavendish there is no need to exaggerate, such as by calling him the founder of water analysis, or by modernizing the language of his thought—he did not give 'a clear exposition of the principle of reciprocal proportions'. The book is well produced but spoilt by misprints and expressions such as 'obtained *approximately* the correct values' (author's italics). F. W. GIBBS

ROYAL SOCIETY

A Brief History of the Royal Society, *by E. N. da C. Andrade. Pp.* 28. *The Royal Society, London.* 1960. 10*s.* 6*d. net.*

In this short book Professor Andrade relates the history of the Royal Society from its earliest 'prehistoric' meetings in London in 1645 and its formal institution on 28th November 1660 to the presidencies of Banks and of Davy, and then to the more recent and more familiar half-century of 'J. J.', Rutherford, Sherrington, and Hinshelwood. There is much about the Society's publications, about the many duties that it discharges, about its independence, about its international outlook, about its rooms and its treasures, and about the many great names that are included in its illustrious company.

Professor Andrade has here composed a lively account of the Royal Society, and of its work and its functions, as well as of its varying fortunes through three centuries, interspersed with much valuable historical information conveyed with sound judgment and occasionally with shrewd comment, all based on a deep scholarly knowledge of his theme. The reproductions are excellent. D. MCKIE

The Royal Society: its Origins and Founders, *edited by Sir Harold Hartley. Pp.* 275. *The Royal Society, London.* 1960. 35*s. net.*

The Tercentenary of the Royal Society in July 1960 was marked by a great gathering in London of representatives of the world's principal centres of scientific research. This memorable occasion prompted a wealth of tributes, both written and spoken, in many of which the origins and history of the Society were authoritatively considered. Such tributes contain much of interest and value to the historian, but naturally special importance attaches to an historical work compiled by the Society itself; this book certainly does justice to a great occasion. As its title indicates, it is not a complete history but, very fittingly, is devoted to the early days. In a most scholarly and elegantly written first chapter Douglas McKie describes the circumstances, intellectual and political, that led to the famous meeting at Gresham College on 28th November 1660, at which the Society was formally constituted. An interesting feature of the chapter is a critical analysis of the justification for identifying Boyle's 'Invisible College' with the group of philosophers from whose informal meetings the Society grew. Although this identification has been accepted for over two centuries there are now considerable doubts about its validity; Boyle may have been referring to some quite different group that cannot now be identified.

This introductory chapter is followed by a series of biographical sketches, twenty-two in all, of the twelve founder members of the Society and some of those who played a leading part in establishing it. Where available, there is a portrait of the subject. They include Charles II (E. S. de Beer); John Wallis (J. F. Scott); William Ball (A. Armitage); Henry Oldenburg (R. K. Bluhm); and Sir Robert Moray (D. C. Martin). In a short epilogue the Editor discusses the Society's contributions to knowledge, to men's ways of thinking, and to the English language.

This book is not only a splendid memento of a great occasion but one of lasting value. Sir Harold Hartley has clearly taken great pains over its preparation and he and his distinguished collaborators deserve high praise for the result.

TREVOR I. WILLIAMS

HISTORY OF TECHNOLOGY

A Short History of Technology from Earliest Times to A.D. 1900, *by T. K. Derry and Trevor I. Williams. Pp. xviii* + 782. *Clarendon Press, Oxford.* 1960. 38*s. net.*

In recent times there has been a strong movement away from the traditional expression of history in political terms, and both social and economic factors have been given more weight. The upheavals of the twentieth century have now laid bare the great importance of technology. It is plain that the lot of man has been revolutionized by the application of science to agriculture, medicine materials, power, and communications. Warfare and power politics have become dominated by technological developments.

It is therefore most important that a history of technology should have appeared in which the quickening rate of these changes can be studied. The book comes from the same stable as the large five-volume work which appeared two years ago (a "History os Technology," edited by Charles Singer, E. J. Holmyard, A. R. Hall, and Trevor I. Williams). Its attractive style and illustrations will please the general reader, and the wealth of information which it contains will be of particular value to students of history and of technology. The great connections between technology and general history are firmly established; but the book is full of charming and human detail, and it never becomes portentous. This excellent work fills a real need.

PATRICK LINSTEAD

Short notices of books

(These notices are descriptive rather than critical and are designed to give a general indication of the nature and scope of the books.)

An Introduction to Linear Programming and the Theory of Games, *by S. Vajda. Pp.* 76. *John Wiley & Sons Inc., New York; Methuen & Co. Ltd, London.* 1960. 9*s.* 6*d. net.*

What is effectively a new branch of science, operational research, has been developed during and since World War II. This book, which is the amplified text of a series of lectures, deals with two mathematical techniques that are fundamental to operational research: there are chapters on the transportation problem; the concept of duality; the fundamental concepts of the theory of games; non-zero-sum games; and more-person games. It is intended to serve as an introduction to the more specialized texts, and it avoids the use of advanced mathematics.

Russian Chemical Reviews. *No.* 1, *edited by J. N. Agar, R. H. Prince, and A. R. Katritzky. Pp.* 51. *The Chemical Society, London.* £12 *net p.a.*

This is a translation of the Russian journal Успехи Химии and is published in monthly parts that correspond to the Russian original, which is edited by V. V. Korshak. The five or six articles in each issue review recent Russian or, occasionally, Chinese work; articles from other East European countries are not translated, nor are those that are translated from other languages. Papers in the first issue of the translation deal, among other topics, with the chemical action of ultrasonic waves on macromolecules (I. E. El'piner), the hydrides of the rare-earth metals (V. I. Mikheeva and M. E. Kost), and the chemistry of quinuclidine derivatives (M. V. Rubtsov, E. E. Mikhlina, and L. N. Yakhontov).

An Introduction to Physical Chemistry for Biologists and Medical Students, *by H. R. Kruyt and J. T. G. Overbeek (translated by A. J. Mee, in collaboration with J. E. Spice). Pp. ix* + 197. *William Heinemann Ltd, London.* 1960. 25*s. net.*

The emphasis of this book is on colloid science, and, broadly speaking, it deals only with those aspects of physical chemistry that are necessary for understanding colloid science. These aspects include osmotic pressure; reaction velocity; and electrokinetic phenomena. The last third of the book is devoted to colloid science itself.

Handbuch der Papierchromatographie, *Vol. II, by I. M. Hais and K. Macek. Pp. xxiv* + 726. *Gustav Fischer Verlag, Jena,* 1960. *DM* 44.

The first volume of this handbook contained the text: this one contains the bibliography and indexes. There are 10 290 references, subdivided into small groups, dealing with paper chromatography as applied to a very wide variety of substances.

Elements of Ion Exchange, *by R. Kunin. Pp. ix* + 164. *Reinhold Publishing Corporation, New York; Chapman and Hall Ltd, London.* 1960. 46*s. net.*

The intention of this book is to provide a basic understanding of ion-exchange techniques for those with little theoretical background in chemistry. The subjects covered include the structure and synthesis of ion-exchange materials, general principles of ion-exchange applications, and ion-exchange in hydrometallurgy and related fields.

The Effects of Pollution on Living Material, *edited by W. B. Yapp. Pp. xii* + 154. *The Institute of Biology, London.* 1959. 25*s. net.*

This, the eighth of the symposia of the Institute of Biology, is the account of a meeting held in London in 1958. Some of the papers are concerned with local problems, such as that of the general condition of the rivers of Britain; but most deal with larger issues, such as the effects of air pollution on plant growth and on man, the uptake of fission products by plants, the relation between pollution and the abundance of animals in estuaries, and fluorosis in farm animals.

Methods of Biochemical Analysis, *Vol. VIII, edited by D. Glick. Pp. ix* + 400. *Interscience Publishers Inc., New York; Interscience Publishers Ltd, London.* 1960. 75*s. net.*

The editor of this series considers that there is a need for reviews concerned with techniques and methods as well as those devoted to progress in the various fields of science. The subjects of the papers in this volume include qualitative and quantitative determination of the fatty acids by gas-liquid chromatography, the measurement of luciferin and luciferase, the isolation and determination of neuraminic acids, and the determination of porphyrins in biological materials.

Plant Marvels in Miniature; a Photographic Study, *by C. Postma. Pp.* 173. *George G. Harrap & Co. Ltd, London.* 1960. 45*s. net.*

The photographs in this book are arranged in series, so as to show, for example, roots, the parts of a spiderwort, and leaf buds, under increasing magnification, up to, in some cases, three thousand times. The photographs are large, and the aim is to provide aesthetically satisfying pictures together with a scientific descriptive text.

Plant Viruses (*third edition*), *by K. M. Smith. Pp. xii* + 209. *Methuen & Co. Ltd, London.* 1960. 16*s.* 6*d. net.*

The first of the two main sections into which this book is divided deals with the basic principles of plant virology and the means by which a knowledge of these principles has been attained. The second part gives a concise account of the practical methods of study, although no attempt is made to give detailed accounts of such specialized and complicated techniques as those necessary in the electron microscopy of viruses. The book is intended for the student with no previous knowledge of plant viruses.

Oysters, *by C. M. Yonge. Pp. xiv* + 209. *Collins, London.* 1960. 21*s. net.*

This book aims at providing a balanced account of oysters, and the chapters cover the structure, the mode of functioning, and the habits of oysters; their use as food; and their cultivation, giving a historical background to the modern techniques.

Fats and Oils (*second edition*), *by H. Kirschenbauer. Pp. vi* + 240. *Reinhold Publishing Corp., New York; Chapman and Hall Ltd, London.* 1960. 56*s. net.*

The author has aimed, when revising the first edition, at keeping the book compact; at the same time he attempts to include the latest information on the chemical structure, methods of processing, and use of fats and oils, both animal and vegetable. Among the subjects covered are autoxidation, rancidity, cholesterol in diet, selective hydrogenation, and sequestering agents.

Notes on contributors

C. G. BUTLER,

M.A., Ph.D.,

Was born in 1913 and was educated at Monkton Combe School, Bath, and Queens' College, Cambridge, where he took a Natural Sciences Tripos. He spent a further five years in Cambridge doing research on whiteflies and locusts, and he was appointed Entomologist in charge of the Bee Section at Rothamsted Experimental Station in January 1939. The Section was made a separate Department in 1944, and he became its first Head. Most of his scientific papers deal with the physiology and behaviour of the honeybee; he is the author of two books and the joint author of a third.

M. K. McQUILLAN,

M.A.,

Was born in 1921 and went to Cambridge in 1939. On graduation in 1942 she joined the Metallurgy Division at the Royal Aircraft Establishment, where her interest in the rarer metals began with titanium. She continued work on titanium both at the Aeronautical Research Laboratories, Melbourne, and at the Metals Division of Imperial Chemical Industries Limited. As leader of the team responsible for research on and small-scale production of 'new' metals, however, her more recent work has been concerned with a wider range of the rarer metals, among which are included those used in the nuclear-energy industry.

T. W. FARTHING,

M.A., Ph.D.,

Was educated at Jesus College, Cambridge, from 1945 until 1951, when he joined the Metals Division of Imperial Chemical Industries Limited. He worked in the Research Department on corrosion and problems associated with copper and brass tube production. He later specialized in the study of 'new' metals and transferred, at its inception, to the project aimed at producing beryllium for use in the atomic-energy industry.

A. L. KURSANOV,

Dr.Bio.Sc.,

Was born in Moscow in 1902 and graduated from Moscow University in 1927 after specializing in plant physiology and biochemistry. His first investigations were performed under the guidance of his father, L. Kursanov, and F. Krasheninnikov, and were on the physiology of diseased plants. He became a Doctor of Biological Sciences in 1940 after defending a thesis on 'The Action of Enzymes in Living Plants'. In 1953 he was elected a Member of the U.S.S.R. Academy of Sciences. He was the head of a laboratory in the A. N. Bakh Institute of Biochemistry from 1940 to 1952. In 1952 he became Director of the K. A. Timiriazev Institute of Plant Physiology of the U.S.S.R. Academy of Sciences in Moscow. His main scientific interests are translocation of organic substances in plants, metabolism in roots in connection with absorption of ions, and biosynthesis and the physiological role of tea catechins.

G. C. BARKER,

M.A., Ph.D.,

Was born in Goole, Yorkshire, in 1915 and entered Cambridge University in 1933; he began research in 1937 on electrochemistry in the Department of Physical Chemistry. During the war years he was mainly engaged on radar work at the Telecommunications Research Establishment, Malvern, Worcestershire, and since 1946 he has been employed by the United Kingdom Atomic Energy Authority, first at Chalk River, Ontario, and later at Harwell. In recent years he has worked on the development of new polarographic methods and on the application of electronic techniques to fundamental electrochemical problems.

G. W. C. MILNER,

D.Sc., F.R.I.C.,

Was born in Doncaster, Yorkshire, in 1918 and was educated at Sheffield University (1936–40). During the war he worked for the Admiralty at the Bragg Laboratory, Sheffield, on research problems connected with the application of physico-chemical techniques to analytical chemistry. After a short period as lecturer in chemistry at Hull University (1946–48), he returned to the Bragg Laboratory and then transferred to Harwell in 1951. At Harwell he has been engaged on problems in analytical chemistry, including applications of the square-wave and pulse polarographs.

T. F. GOREAU,

Ph.D.,

Was born in 1924 and was educated at Clark University and Yale University. He is now Senior Lecturer in Physiology at the University College of the West Indies, Jamaica, where he has been teaching since 1951. He is also associated with the Department of Marine Ecology and Biochemistry of the New York Zoological Society. His interest in research is in problems of skeletogenesis in reef-building corals. He is carrying out underwater ecological studies on community structure in West Indian reefs as well as laboratory investigations, and is working on correlations between physiological and environmental factors influencing calcification in corals.

E. GLUECKAUF,

Dr.-Ing., M.Sc., D.Sc.,

Was born in 1906 and studied chemistry at the University, and at the *Technische Hochschule*, in Berlin. Subsequently he did research with F. A. Paneth, first at Imperial College, London, and then at Durham. In 1947 he joined the Atomic Energy Research Establishment at Harwell, where he built up the Physical Chemistry Group. His main interests in chemistry are in the theory of chromatography and ion exchange, in microconstituents of the atmosphere, and in specialized fields of radioactivity. He is now a Deputy Chief Scientist at Harwell, in charge of the Radio Chemistry Branch.

Jubilee of the *Max-Planck-Gesellschaft*

This year marks an important anniversary in German science, the jubilee of the foundation of the *Kaiser Wilhelm-Gesellschaft zur Förderung der Wissenschaften.* Since the foundation of this Institution was in itself directly linked with the celebration of the centenary of the University of Berlin, the event is a doubly significant one in the history of science in Germany.

The foundation of Berlin University was a consequence of earlier political developments. When Napoleon created the Kingdom of Westphalia for his brother Jerome in 1807, Prussia lost the famous University of Halle. To replace it, Frederick William III set up in 1810 a new university to provide a centre for learning in his capital. Today this university lies in the East Sector of Berlin, and has been renamed the Humboldt University after Karl Wilhelm von Humboldt. The Free University of Berlin was established in 1948 to serve the needs of the West Sector.

Shortly before the celebration of the centenary of Berlin University, Adolf von Harnack, the great theologian, prepared for the Kaiser a concise but forcefully worded memorandum stating the case for the establishment in Germany of centres for pure research. Although Humboldt himself had stressed the need for original work a century earlier, the emphasis in the German universities at that time was on teaching rather than research. Van't Hoff, who had come to Berlin from Amsterdam in 1896, was almost unique in holding a professorship that entailed no teaching duties. Von Harnack's memorandum had the effect desired, for in the course of the Berlin University centenary celebrations the Kaiser announced a scheme for the setting up of the group of research institutes that in due course became known as the *Kaiser Wilhelm-Gesellschaft.* The necessary endowment was raised by public subscription and Adolf von Harnack became the first President. After the first world war, support from the public and from industry proved insufficient, and the Institution was increasingly supported by the state.

On von Harnack's death in 1930 the Presidency was assumed by Max Planck, who, after an already brilliant career that included the award of a Nobel Prize in 1918, had joined the board of management of the physico-chemical laboratory in 1926; that it was able to attract so great a scientist is a measure of the reputation that the *Kaiser Wilhelm-Gesellschaft* had established in a surprisingly short time. He became the head of a research organization of very wide ramifications, carrying out work in all the principal fields of chemistry, biology, and medicine; in all there were by then some thirty separate research institutes in various parts of Germany, with a few situated in other countries. In Berlin, for example, were the institutes for chemistry; for physical chemistry and electrochemistry; and for experimental medicine. Research relating to iron and steel was carried out at Düsseldorf, and on problems relating to coal and its utilization at Mülheim. Outside Germany were a marine biological research laboratory in Italy and a microbiological laboratory in Brazil.

One of the aims of the *Kaiser Wilhelm-Gesellschaft* was the creation of an institute devoted to pure physics, as distinct from applied physics, which was already the province of the *Physikalisch-Technische Reichsanstalt*—a model for Britain's National Physical Laboratory of 1900—that had been established at Charlottenburg in 1888, with the great von Helmholtz as its first director. Although Einstein was proposed for, and accepted, the post of director, the physics institute was not in fact built until the 1930s, owing to the first world war and the difficulties that followed it. As long as the institute had no building of its own, its activities were, in the main, restricted to the award of research fellowships tenable at other seats of learning. In the desire to attract men of the very highest calibre, special inducements were offered to Einstein to persuade him to return to Germany from Switzerland. He was made a member of the *Preussische Akademie der Wissenschaften*, and given a professorship in Berlin University that left him free to pursue his research. Further, he was allowed the exceptional privilege of retaining his Swiss nationality.

Commemorative volumes published when the *Kaiser Wilhelm-Gesellschaft* celebrated its twenty-fifth anniversary in 1936 remind us that associated with it had been a remarkably large number of men of the highest international repute in science and technology, of whom only a few can be mentioned here. At the physico-chemical and electrochemical laboratory, for example—situated in what was named, as a courteous gesture to British science, the *Faradayweg*—the board of management included not only Planck, but Carl Bosch, a Nobel Laureate of 1931. The name of Bosch will always be remembered with that of Fritz Haber, awarded a Nobel Prize for chemistry in 1918, for the Haber-Bosch process of ammonia synthesis has long been, and shows every sign of remaining, one of the most important of all industrial chemical processes. Haber, who died in 1934, had been associated with this institute from its earliest days, having been appointed its first director in 1911. At the Institute for Chemistry was Otto Hahn, whose Nobel Prize was to be awarded in 1944; closely associated with Hahn was Lise Meitner. Yet another Nobel Laureate, Otto Warburg, was at the Institute for Cytophysiology. The names of Gustav Krupp, Karl von Siemens, Albert Vögler, and Fritz Thyssen on the board of management that controlled the Institution are reminders of the great interest of German industrialists in the project.

Unhappily, however, the twenty-fifth anniversary celebrations corresponded with a serious decline in the fortunes of the *Kaiser Wilhelm-Gesellschaft*. 1933 had seen the National Socialists swept into power, with all the dire consequences that this event entailed for Germany and the world. The violent anti-semitism of the new régime was tempered by neither sense nor humanity; not even such men as Einstein were spared, despite the exceptional concessions that had been made to induce him to return to Germany in the first instance. When the persecution began, those Jewish scientists fortunate enough to escape with their lives found refuge in other countries, where many were able to resume their researches. Germany's loss was the world's gain.

This unhappy period, culminating in six years of war, is one best forgotten in the present context. Its end found the *Kaiser Wilhelm-Gesellschaft*, in common with virtually all other German scientific institutions, utterly disorganized and the staffs of the individual institutes widely scattered. Very early in the allied occupation, the British authorities gave much thought to how academic science in Western Germany could best be revived, and inevitably the future of the *Kaiser Wilhelm-Gesellschaft* received particular attention. Immediate steps were taken—through a scientific advisory council in Göttingen, composed of leading professors from the university there—to re-establish contact between individual laboratories of the Institution in the British and, where possible, other zones. In the new circumstances a change of name seemed desirable, and after careful deliberation it was decided to commemorate the life and work of Max Planck. The British occupation authorities were the first formally to recognise the *Max-Planck-Gesellschaft*, with Otto Hahn as its President; a year was to pass before the American and French governments followed suit.

The change of name, marking the beginning of a new era, was a most appropriate one. Max Planck, President of the Institution in happier days was internationally held in respect and admiration for both his scientific genius and his personal qualities. The quantum theory which he formulated at the beginning of this century can fairly be said to dominate the whole field of modern theoretical physics, and it is unnecessary in these pages to enlarge upon its importance. In 1946 Planck came as an honoured guest to England; although great age had brought physical infirmity, his mind was still extraordinarily clear. The occasion was a particularly appropriate one, for it was the Royal Society's international celebrations to mark the tercentenary of Newton's birth. Newton, it will be recalled, ascribed to light certain wave properties, yet it seemed to him that in some respects the corpuscular theory was the more satisfactory. Planck's work both emphasized the dual nature of radiation and paved the way to later research that finally, through wave mechanics, resolved the paradox and established a unity of wave and particle.

Today, the *Max-Planck-Gesellschaft* stands firmly established. In its more than forty institutes important contributions are being made to biology, chemistry, medicine, and physics. In addition, there are three institutes for the study of law and one for economics; in Rome is the Hertz Library, now the Institution's only establishment outside Western Germany. Some of these laboratories represent the rebuilding of ones damaged or destroyed during the war, others are entirely new. Although financial support comes mainly from the state governments, with some industrial and private help, the independence of action that was so important a feature of the early years of the *Kaiser Wilhelm-Gesellschaft* has been regained. There are firm links with the universities, and several directors of institutes are also university professors. The high standard of the early years seems to have been fully regained, for the staff again includes several Nobel Laureates.

This regaining of its former lustre is typical of German science generally. ENDEAVOUR is happy to send congratulations and good wishes for the future to the *Max-Planck-Gesellschaft* on this important anniversary in its history.

Electron-deficient compounds

F. G. A. STONE

A fairly large group of covalent compounds is known that does not conform with the usual rules relating chemical composition with classical valence theory. These substances, in which there appear to be fewer valence electron-pairs than there are chemical bonds, have been described by the term 'electron-deficient'. This article describes the modern view of the structure and bonding of these compounds, as well as some of their distinctive chemical reactions.

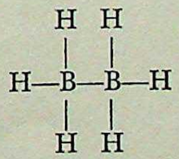
In the century that has elapsed since the idea of chemical structure—in the sense of definite linkages between atoms—began to be accepted, our understanding has increased to such a point that knowledge of the physical state of a substance and of the geometrical arrangement of the atoms in it enables us to give a fairly adequate description of the nature of the chemical bonds present in the substance. The partial dispersion of the veil of mystery concealing the nature of chemical bonds has been due to the efforts of many scientists. Among early workers the names of Butlerov [1], Kolbe, Frankland, and Kekulé come quickly to mind. Among the pioneers the name of G. N. Lewis [2] occupies a special position, for it was he who laid the foundations of the electronic theory of valence, recognizing that stable shells of electrons could be created by the transfer of electrons from one atom to another, forming ions, or by the sharing of two electrons between two atoms, forming a covalent bond. These ideas were extended and sharpened by others, notably I. Langmuir, N. V. Sidgwick, and L. Pauling.

For the majority of covalent substances the electron-pair theory of valence, refined by the application of quantum mechanics, and underlain by the assumption that each bond involves two atoms only, has reasonable validity. However, contemporary with the birth of the electronic theory of valence was the identification of a class of covalent substances that did not fit within the framework of the electron-pair theory of valence. These substances were a series of volatile boron hydrides [3].

Because boron is in Group III of the periodic classification of the elements, the formula of its simplest molecular hydride should be BH_3 (borane); instead, the simplest molecular hydride has the dimeric formula B_2H_6 (diborane). The properties of diborane, such as the boiling point of $-92{\cdot}5$° C, are in accord with the presence in the hydride of essentially covalent bonds. However, a structure in which all the linkages are conventional electron-pair bonds requires at least fourteen valence electrons to hold the hydride together. In diborane the number of valence electrons is only twelve. The molecule is therefore said to be 'electron-deficient', this term being used to describe any compound having too few valence electrons to permit all atoms to be held together by two-atom electron-pair bonds.

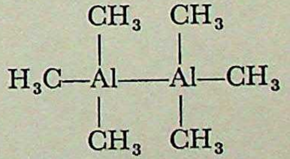
The property of electron-deficiency is not limited to the hydrides of boron. Discovery that trimethylaluminium is dimeric [4] and that tetramethylplatinum is tetrameric [5] led to the realization that a few substances having metal–carbon bonds were also electron-deficient. As with diborane, the polymeric character of trimethylaluminium and tetramethylplatinum implies the existence in these substances of an excess of bonds over electron-pairs.

For many years the study of electron-deficient compounds was hampered by the fact that chemists were led into attempts to explain geometrical structures that were, in fact, incorrect. Thus, the electron-diffraction technique, less reliable than at present, gave results which appeared to favour an ethane-like structure (I) for diborane, and a structure (II) for the dimer of trimethylaluminium like that of hexamethylethane:



$$\begin{array}{ccc} & H\quad H & \\ & |\quad\;\; | & \\ H- & B-B & -H \\ & |\quad\;\; | & \\ & H\quad H & \end{array}$$

I



$$\begin{array}{ccc} & CH_3\quad CH_3 & \\ & |\qquad\quad | & \\ H_3C- & Al-Al & -CH_3 \\ & |\qquad\quad | & \\ & CH_3\quad CH_3 & \end{array}$$

II

The correct configuration of the atoms in diborane (figure 1) was first established by a study [6] of its infra-red spectrum, and subsequently confirmed by a variety of other physical evidence [7]. The dimer of trimethylaluminium has a

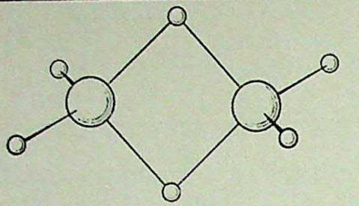


FIGURE 1 – *The atomic arrangement in diborane* (B_2H_6).

structure [8] analogous to that of diborane, with methyl groups instead of hydrogen atoms, and aluminium atoms instead of boron atoms. Establishment of the dispositions of the atoms in electron-deficient compounds is of paramount importance, since without knowledge of geometrical structures any discussion of the distribution of electrons between atoms is useless. It is fortunate, therefore, that during the last decade rapid progress in the determination of the structures of electron-deficient compounds has been made by the X-ray diffraction method, principally by W. N. Lipscomb [9] and R. E. Rundle [10] and their respective co-workers.

The feature common to all electron-deficient compounds is an atom, usually of a metal, having fewer valence electrons than stable bonding orbitals, and occurring in combination with atoms or groups containing no unshared electron-pairs. In these circumstances, classical valence theory, with its assumption that a covalent bond is a linkage localized between two atoms only, is quite unsatisfactory. In an electron-deficient compound, delocalization of the electrons occurs so as to employ all of the low-energy orbitals of the metallic atom in bonding. This situation leads to a relaxation of the stereochemical restrictions that govern the disposition of atoms in electronically saturated molecules. These geometrical restrictions are imposed because, in atoms with all valence orbitals filled, the different electron-pairs tend to avoid one another as far as is possible. Thus, as has been pointed out by Pauling [11], in most electron-deficient compounds the metal atom has a ligancy (co-ordination number) not only greater than the number of valence electrons, but also greater than the number of low-energy orbitals. Moreover, atoms adjacent to an electron-deficient atom increase their co-ordination number to a value above that of their orbital number.

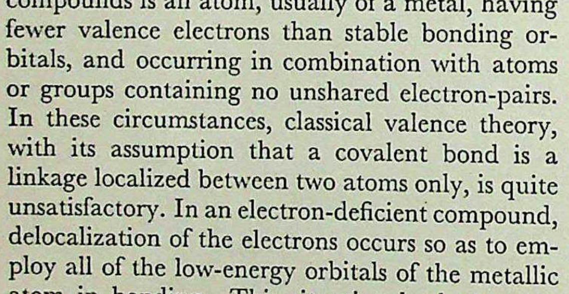


FIGURE 2 – *The structure of beryllium chloride.*

These principles may be illustrated by a consideration of the structures of beryllium chloride and dimethylberyllium, the latter, but not the former, being an electron-deficient compound. In many beryllium compounds, the beryllium atoms form four bonds that tend to be directed towards the corners of a tetrahedron, as in solid beryllium chloride, a chain polymer (figure 2) [10]. A beryllium atom in the gas phase in its lowest energy-state has four electrons, two being paired in the 1s orbital and two being paired in the next most stable orbital, the 2s, so that the electronic configuration of the atom is described as $1s^2 2s^2$. This description of the electron distribution in a free beryllium atom, involving a filled inner shell and a filled sub-shell, is clearly of no help in explaining the structure of beryllium chloride. Difficulties of this type have been circumvented by the application of quantum-mechanical principles and the concept of hybridization of bond orbitals [11]. For an electron associated with an atom, three 2p atomic orbitals correspond to the next energy level above the 2s orbital. It is found that a linear combination of these four orbitals yields bond orbitals of lower energy than those obtainable from s or p orbitals alone. The four so-called hybrid orbitals, the maximum number permissible in the valence shell of beryllium, are disposed tetrahedrally in space. However, a beryllium atom, possessing only two valence electrons, requires six more electrons from other atoms or ions to attain the especially stable arrangement of a completed electron-shell. In a monomeric beryllium chloride molecule, the valence shell of the beryllium atom, which must share its electrons with two chlorine atoms, contains only four electrons. In the gas phase, at elevated temperatures, it is possible for the monomeric species $BeCl_2$ to exist. In this event each beryllium–chlorine linkage can be regarded as an electron-pair bond, represented as being formed by overlap of a beryllium sp-hybrid orbital with a p-orbital of one of the chlorine atoms. Now a chlorine atom forming a single bond with another atom possesses non-bonding electron-pairs in p-orbitals. These electron pairs can be used to link other atoms, provided the latter have vacant orbitals available. The linkage formed in this

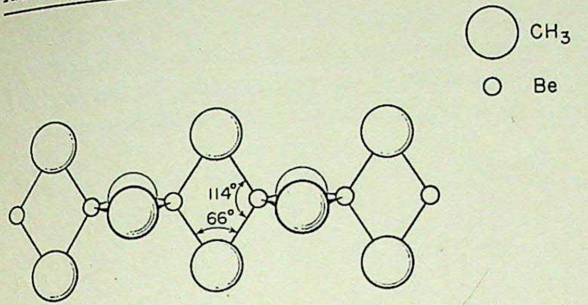
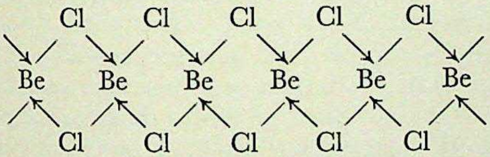


FIGURE 3 – *The structure of dimethylberyllium.*

manner is a dative bond, with both the electrons necessary for bond formation supplied by one atom. Since in monomeric beryllium chloride the beryllium atom does not have a completed shell, and since beryllium can have a total of four valence-orbitals, it is not surprising that, under normal temperature conditions, the monomeric species $BeCl_2$ polymerizes (III) so as to use all low-energy (valence) orbitals of beryllium, giving both beryllium and chlorine completed electron-shells. The resultant geometrical structure[1] is perforce a compromise between the four beryllium orbitals, which ideally would be tetrahedrally disposed, and the two chlorine orbitals, which ideally would be at right angles. Halogen bridges of the type found in beryllium chloride are fairly common in the metal halides. Thus the trichlorides of aluminium and indium are dimers in the gas phase, and the palladium chloride $PdCl_2$ is a polymeric solid. It is important to note that these polymeric halides are not electron-deficient, since every metal-halogen bond contains an electron-pair. Although the metals have more orbitals than valence electrons, electron-deficiency does not occur, because the halogen atoms bonded to the metals are able to contribute additional electrons.



III

The situation is very different in solid dimethylberyllium (figure 3), of which the vapour pressure is 0·6 mm at 100° C. In this compound, methyl groups serve as bridges between beryllium atoms [10]. Since methyl groups, unlike chlorine atoms, have no electron-pairs available for donation, the beryllium-carbon bonds cannot be treated as localized electron-pair bonds, and dimethylberyllium is electron-deficient. The co-ordination number of carbon has been raised to five, a value unusual for carbon, and, although the nearest neighbours of each beryllium atom are four methyl groups, the acute-angled group, Be—C—Be, has allowed adjacent beryllium atoms to approach closely enough to suggest some metal–metal interaction and also a ligancy of six for beryllium.

BONDING IN ELECTRON-DEFICIENT COMPOUNDS

The distribution of electrons in electron-deficient compounds has been the subject of much speculation for almost fifty years. Only during the last decade, however, has a reasonably satisfactory rationale of the bonding in these compounds been given. The modern viewpoint is based on a consideration of molecular orbitals rather than atomic orbitals.

A fundamental principle of the molecular-orbital theory of valence, as applied to a diatomic molecule AB, is that two molecular orbitals can be derived by a linear combination of an atomic orbital from atom A with an atomic orbital from atom B.[2] Of the two resultant molecular orbitals, one is of lower energy (bonding) and one of higher energy (antibonding) than either of the two original atomic orbitals (figure 4*a*). The electron-pair will occupy the orbital of lowest energy, the bonding orbital. Similarly, when three atomic orbitals on three centres are able to overlap, one obtains three molecular orbitals of which one

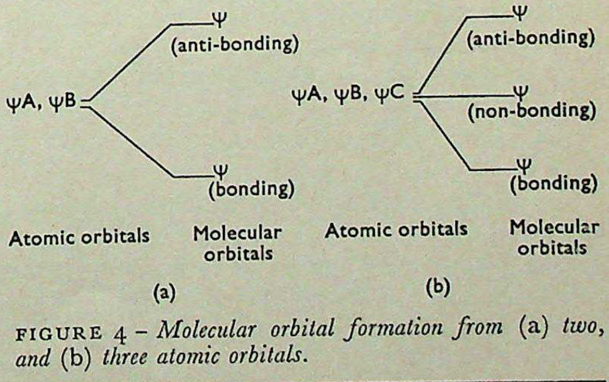


FIGURE 4 – *Molecular orbital formation from* (a) *two, and* (b) *three atomic orbitals.*

---

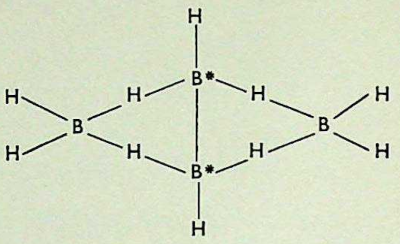
[1] Although the formula as given represents some bonds as donor bonds and others as ordinary covalent bonds, it should be recognized that it is not in fact possible to distinguish between the bonds.

[2] The approximation of the linear combination of atomic orbitals is applicable if the two atomic orbitals on the two atoms have approximately the same energies and the same symmetry relative to the molecular axis AB, and the charge clouds overlap sufficiently.

only is strongly bonding (figure 4*b*). Again there is only one state of lowest energy, and for this ground state only a single electron-pair is required, although three atoms are involved.

We are now in a position to see how the molecular-orbital (MO) correlation diagram for a three-centre bond [12] (figure 4*b*) can be applied to the electron-deficient B_2H_6 molecule [9, 13]. The 2s and three 2p atomic orbitals of the two boron atoms are hybridized to give each boron atom a set of four equivalent tetrahedrally disposed $2sp^3$-orbitals. Each boron atom uses two of its sp^3-hybrid orbitals to attach two hydrogen atoms. These are the four terminal hydrogen atoms, and the B—H bonds are of the normal electron-pair type, a fact suggested by force-constant measurements. Two sp^3-orbitals remain on each boron atom, and each bridging hydrogen atom possesses a 1s-orbital, so that six atomic orbitals have still to be considered. For the filling of these orbitals, four valence-electrons remain, eight having been used in the terminal B—H links. The B—H—B bridges may then be described in terms of two localized molecular orbitals, each formed by combination of two sp^3 hybrid atomic orbitals on the two boron atoms and a 1s-orbital of hydrogen. Since a linear combination of atomic orbitals from three atoms leads to a unique lowest-energy state (figure 4*b*), the four valence electrons, divided into two pairs, are sufficient to fill the two 'three-centre' bonds that are concentrated on either side of the boron-boron axis (figure 5*a*). The polyatomic diborane molecule can thus be described in terms of filled valence-orbitals by invoking the concept of two localized molecular-orbitals extending over three atomic nuclei. Clearly, such three-centre bonds involve more electron delocalization than the normal two-centre bond.

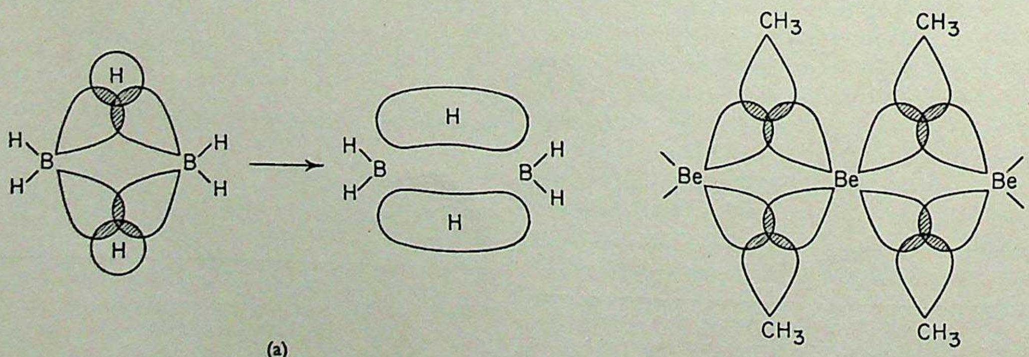
Three-centre bonds can also be used to describe the bonding in some of the higher boron hydrides. Consider, for example, tetraborane (figure 6), again assuming sp^3-hybridization for boron. The hydride B_4H_{10} has twenty-two valence electrons.



IV

Six hydrogen atoms are bonded to boron atoms by six electron pairs, and the four B—H—B three-centre bonds (IV) require four electron pairs. These ten pairs of electrons fill all the hydrogen orbitals and all the boron orbitals, except for one orbital on each of the two boron atoms B* (IV). There remains one electron pair for these two boron orbitals that are directed towards one another and can therefore form an electron-pair B—B bond. The same bonding principle can be used for the hydride B_5H_{11}, except that it is necessary to consider two three-centre bonds of the type B—B—B as well as three bonds of the type B—H—B. It is interesting to note the diborane type of geometry that occurs in the open parts of the higher boron hydrides (figure 6).

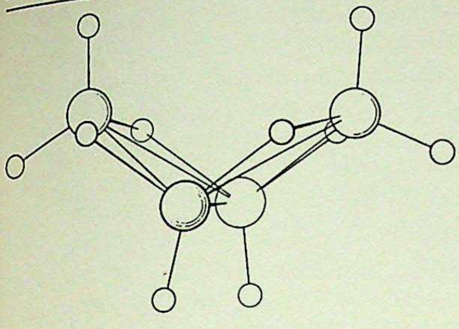
Molecular orbitals localized in the sense that three atom-centres are considered separately from other atoms in the molecule are also a useful concept for describing the bridge bonding in the trimethylaluminium dimer and polymeric dimethylberyllium (figure 5*b*). A single stable localized molecular orbital is obtained by combining an sp^3-hybrid carbon orbital with sp^3-orbitals of adjacent aluminium or beryllium atoms. An



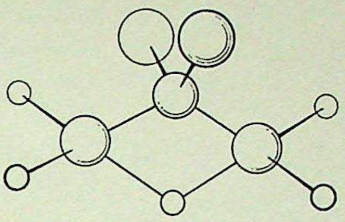
(a)

(b)

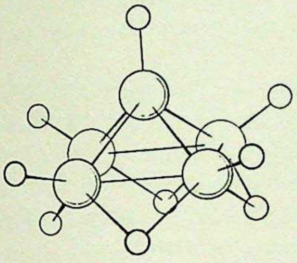
FIGURE 5 – (a) *Bridge bonding in diborane.* (b) *Bridge bonding in polymeric dimethylberyllium.*



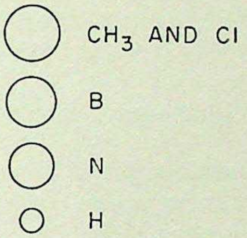
The structure of B_4H_{10}.

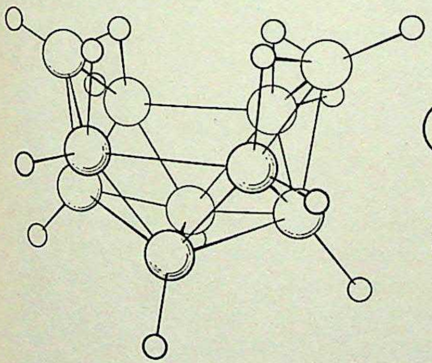


The structure of $(CH_3)_2NB_2H_5$.

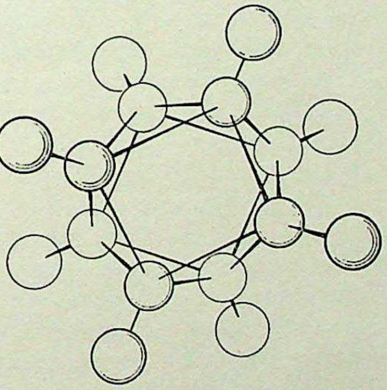


The structure of B_5H_9.





The structure of $B_{10}H_{14}$.



The structure of B_8Cl_8.

FIGURE 6 – *Electron-deficient boron compounds.*

electron-pair is then available for the filling of the three-centre bonds.

In electron-deficient compounds that are more complex than those mentioned above, greater electron delocalizations occur, and the three-centre-bond concept has to be abandoned in favour of multicentre bonds. Thus in tetramethylplatinum (figure 7) a four-centre orbital is required to bond a methyl group to three platinum atoms using one electron pair. Similarly, in the pentaborane B_5H_9 (figure 6) it is necessary to assume a five-centre orbital, involving the five boron atoms in addition to three-centre and other orbitals, in order to obtain a satisfactory description of the molecule.

R. E. Rundle has pointed out [10] that the electron delocalization that is characteristic of electron deficiency increases in a continuous manner as the ratio of valence orbitals to valence electron-pairs increases. One example of this tendency is afforded by the hydrides of boron and aluminium. Like its Group III congener boron, aluminium possesses three valence electrons. However, elements in the row below boron in the Periodic Table are not limited to four valence orbitals but may form up to six. Thus it is not surprising that whereas the BH_3 group dimerizes to B_2H_6, the AlH_3 group displays greater electron-deficiency, forming a highly polymeric solid $(AlH_3)_n$. Similarly, lithium alkyls are more polymeric and therefore more electron-deficient than trimethylaluminium. The greater electron-deficiency of covalent lithium compounds is to be expected, because in a covalent compound, with the implied principle of electron sharing, lithium has but one valence electron to contribute to a total of four possible orbitals.

Metals themselves can be regarded as highly electron-deficient. There are so many orbitals and so few electrons that the electron-pair theory of valence with its localized bonds has little meaning, and other theories have to be advanced to account for metallic bonds.

FURTHER EXAMPLES OF ELECTRON-DEFICIENT COMPOUNDS

To emphasize the widespread occurrence of electron-deficiency, some further examples may be cited.

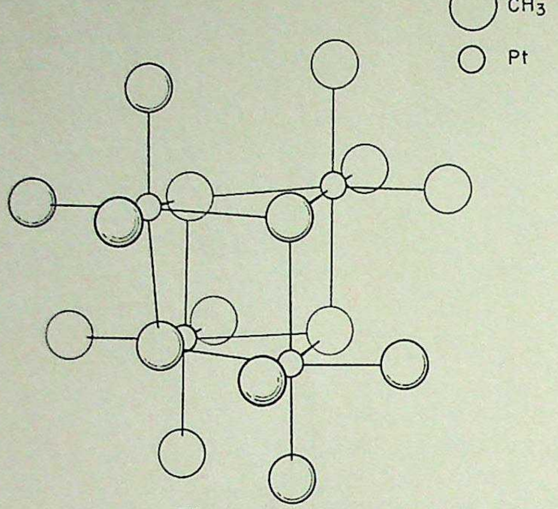


FIGURE 7 – *The structure of tetramethylplatinum.*

Nearly all the known boron hydrides form derivatives in which one or more hydrogen atoms are replaced by other atoms or groups [14]. Bromodiborane (B_2H_5Br), for example, has a structure analogous to diborane, with one of the terminal hydrogen atoms replaced by a bromine atom. Numerous substitution derivatives of decaborane (figure 6) are also known, such as $B_{10}H_{12}(NCCH_3)_2$ and $B_{10}H_{13}I$ (two isomers reported). Substituted boron hydrides of this type contain B—H—B bridges and are electron-deficient, just as are ions derived from the hydrides (such as $B_{10}H_{13}^-$ and $B_3H_8^-$), and certain compounds (such as $(CH_3)_3N.B_3H_7$) obtained by treating boron hydrides with electron-pair donors.

There are many other substances having metal–hydrogen–metal bridges, among them the covalent members of the class of compounds called borohydrides that has been studied over the past 20 years by H. I. Schlesinger and his co-workers. The covalent borohydrides, such as $Be(BH_4)_2$, $Al(BH_4)_3$, $Ti(BH_4)_3$, and $U(BH_4)_4$, have boron–hydrogen–metal linkages, and, for substances containing metals, are exceedingly volatile. Aluminium borohydride, for example, boils at 44·5°C.

Related to the covalent borohydrides, in the sense that they have hydrogen-bridge bonds, are certain polymeric hydrides. Aluminium hydride, mentioned earlier, is representative of this class of hydride, and other examples include hydrides of beryllium, zinc, and cadmium. These polymeric hydrides, with their large degree of cross-linking of metals through hydrogen, are highly electron-deficient. On the other hand, substances that are electron-deficient only to a limited extent are exemplified by the thiodiboranes (RSB_2H_5) and the aminodiboranes ($R_2NB_2H_5$) described by A. B. Burg and his co-workers [16]. In these compounds, such as dimethylaminodiborane [$(CH_3)_2NB_2H_5$] figure 6, only one electron-pair three-centre B—H—B bond is present. The B—S—B and B—N—B bridges are not electron-deficient, since electron-pairs on sulphur and nitrogen can donate to adjacent boron atoms, just as lone pairs on chlorine donate to beryllium in polymeric beryllium chloride (III). In other cases, the occurrence of electron-deficiency depends on the physical state. Thus trimethylindium (m.p. 88·4°) is monomeric as vapour but polymeric as a solid [17].

An interesting example of electron-deficiency is provided by the boron halides B_4Cl_4 and B_8Cl_8 [9], which are highly condensed molecules so that, as mentioned earlier, use of three-centre orbitals is not satisfactory, and a description based on multicentre molecular orbitals is to be preferred. The boron atoms in B_4Cl_4 are linked together tetrahedrally, and each is bonded to a chlorine atom lying outside the tetrahedron. The structure of B_8Cl_8 involves an interesting polyhedron of boron atoms (figure 6). In B_4Cl_4 the B—Cl bonds may as a first approximation be regarded as electron-pair bonds, leaving eight valence electrons available to hold the four boron atoms together. It is possible to show [13] that the atomic orbitals remaining on the tetrahedrally arranged boron atoms can be compounded into six bonding molecular orbitals. The first four of these orbitals employ the eight available valence electrons, while the remaining two have the correct symmetry to accept electrons from the filled p-orbitals possessed by the chlorine atoms. A filled-orbital description of the halide is thus possible. A somewhat similar description can be given to the bonding in B_8Cl_8 [9] and to the bonding in certain electron-deficient borides that contain octahedra of boron atoms, such as CaB_6 [13].

CHEMICAL REACTIVITY

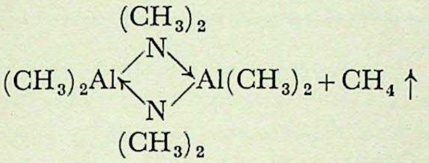
In general, electron-deficient compounds react readily with substances able to donate electrons, the 'nucleophilic reagents' or 'Lewis bases'. Such reactions can give a wide variety of products, and many hundreds of interesting compounds have been obtained from electron-deficient substances in this manner. This short article can, of necessity, list only a few reactions of this kind.

Trimethylaluminium and dimethylberyllium are depolymerized by amines and phosphines, forming addition compounds

$$(CH_3)_3N{\rightarrow}Be(CH_3)_2 \quad \text{or} \quad (CH_3)_3P{\rightarrow}Al(CH_3)_3$$

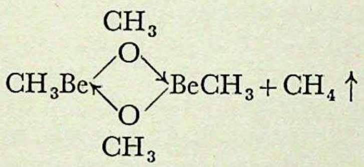
in which all bonds are of the electron-pair type. If the donor atom carries a hydrogen atom, the initial reaction product releases methane on heating:

$$(CH_3)_2NH.Al(CH_3)_3 \xrightarrow{\text{heat}}$$



Sometimes treatment of an electron-deficient compound with a Lewis base leads to so rapid and vigorous a reaction that no intermediate co-ordination compound is detected:

$$[(CH_3)_2Be]_n + CH_3OH \longrightarrow$$



The electron-deficient metal alkyls react with water with explosive violence, probably through intermediates such as $H_2O{\rightarrow}Al(CH_3)_3$.

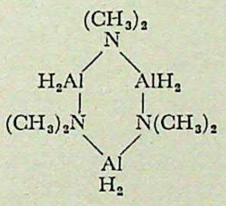
The chemistry of the boron hydrides has received extensive study [14], and most reactions may be assigned to two main classes. Either a borane group may be removed from a boron hydride structure, as in

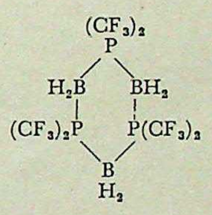
$$B_2H_6 + 2CO \xrightarrow{\text{pressure}} 2OC.BH_3$$

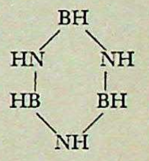
or a BH_2^+ group is removed, as in

$$B_4H_{10} + 2NH_3 \xrightarrow{\text{low temperature}} [H_2B(NH_3)_2]^+(B_3H_8)^-$$

Reactions of boron hydrides involving removal of BH_3 groups are so common that there are some 50 known addition compounds of the type $L.BH_3$, where L is an electron-pair donor. A few analogous complexes of the AlH_3 group are also known, such as $(CH_3)_3N{\rightarrow}AlH_3$. Unlike the borane compounds, the AlH_3 co-ordination compounds are associated to some extent in organic solvents. Association probably involves Al—H—Al bridges, with aluminium using more than four orbitals (see above), so that the compounds are still electron-deficient to some extent. Frequently an electron-donor molecule will form an AlH_3 or BH_3 addition compound that will subsequently decompose, under appropriate conditions, with release of hydrogen. The following are representative of a large number of compounds made in this way.







REFERENCES

[1] Butlerov, A. M. *Z. Chem. Pharm.*, **4**, 549, 1861.

[2] Lewis, G. N. *J. Amer. chem. Soc.*, **38**, 762, 1916.

[3] Stock, A. 'Hydrides of Boron and Silicon'. Cornell University Press, Ithaca, New York. 1933.

[4] Laubengayer, A. W. and Gilliam, W. F. *J. Amer. chem. Soc.*, **63**, 477, 1941.

[5] Rundle, R. E. and Sturdivant, J. H. *Ibid.*, **69**, 1561, 1947.

[6] Price, W. C. *J. chem. Phys.*, **16**, 894, 1948.

[7] Stone, F. G. A. *Quart. Rev. chem. Soc., Lond.*, **9**, 174, 1955.

[8] Lewis, P. H. and Rundle, R. E. *J. chem. Phys.*, **21**, 986, 1953.

[9] Lipscomb, W. N. *Advanc. inorg. Chem. Radiochem.*, **1**, 146, 1959.

[10] Rundle, R. E. *J. phys. Chem.*, **61**, 45, 1957.

[11] Pauling, L. 'The Nature of the Chemical Bond' (Third edition). P. 363. Cornell University Press, Ithaca, New York. 1960.

[12] Longuet-Higgins, H. C. *J. chim. phys.*, **46**, 275, 1949.

[13] *Idem. Quart. Rev. chem. Soc., Lond.*, **11**, 121, 1957.

[14] Stone, F. G. A. *Advanc. inorg. Chem. Radiochem.*, **2**, 279, 1960.

[15] Schlesinger, H. I., Brown, H. C. *et al. J. Amer. chem. Soc.*, **75**, 186, 1953.

[16] Burg, A. B. *Rec. Chem. Progr.*, **15**, 159, 1954.

[17] Amma, E. L. and Rundle, R. E. *J. Amer. chem. Soc.*, **80**, 4141, 1958.

Space research

SIR HARRIE MASSEY

Because of both its very high cost in terms of money and skilled manpower, and the uncertainty about its potential value, space research has inevitably excited a good deal of controversy. This article considers what is meant by space research, what it entails, and what we may reasonably hope to learn from it. Particular attention is paid to distinguishing the roles of the two principal space vehicles, vertical sounding rockets and satellites.

THE MEANING OF SPACE RESEARCH

The availability of rockets with high-power motors, and of accurate methods of guiding, controlling, and tracking them, has far-reaching consequences, but there is at present much confusion about the meaning of the term 'space research'. To many, the first meaning that springs to mind is space travel—voyaging by man in space, including landings on the Moon and planets. Such journeys have so romantic an appeal, representing such an outstanding challenge to man's ingenuity, endurance, and adaptability, that many consider no other justification to be necessary. Other proponents of manned space-travel suggest great advantages to be gained from it, ranging from the colonizing of celestial objects, thereby relieving the problem of overpopulation, to the possibility of tapping new sources of mineral and other wealth. Still others would say that through manned travel we can add greatly to our knowledge of the universe around us and particularly about our nearest neighbours, the Moon, Venus, and Mars. They point out that one of the most fascinating problems of all—to what extent life is confined to this planet—might well find an answer if Mars or Venus could be explored. Such knowledge would not only be of the greatest interest for its own sake but might have applications whose consequences are quite unforeseeable. We must remember, however, that much research may be carried out automatically by instruments carried in a space vehicle, making measurements and transmitting the results back to Earth by means of coded radio signals. So much can be done in this way that it is at least questionable whether the great additional effort and expense necessary to put man safely into space would be justified for scientific observation alone. This does not in itself imply, of course, that the present effort being made to realize manned space-travel is unjustified, for this involves other considerations than scientific research.

It is the scientific research made possible by the availability of space vehicles that the writer prefers to call space research, but he is well aware that this is a purely personal choice and that the term is often employed in a much wider sense. For this reason there has been a tendency to use the new term 'space science' to apply only to the scientific research, but this also is gradually being extended in meaning to cover the whole field. In this article we shall stick to the narrower meaning and discuss no further the case for man in space. On the other hand, the new scope for scientific research will be considered in some detail, with reference also to commercial applications.

In advocating such space research we must be clear that the new knowledge to be gained by the use of automatic equipment carried in space vehicles is of wide range and importance. It is of great value, for example, to the astronomy of the Sun and stars; the physics of the Earth and planets; the study of solar-terrestrial relations; meteorology; the study of cosmic rays; and biophysics. Already there are possible applications of commercial importance, especially to world-wide radio communications and navigation, and as the work develops, there will doubtless be many more.

Before proceeding, one other point of confusion needs to be cleared up. This is one that occurs within the purely scientific part of the subject, and arises through failure to appreciate that space research, as we understand it, is a unity of technique, not of scientific discipline. It is not considered, for example, that radioastronomy, or indeed optical astronomy generally, is part of space research; only those parts of the subject that depend on observations made with equipment in space vehicles are so considered. The realization of this situation is important, as it would be most undesirable for the programme of space research to develop outside the scientific disciplines concerned. They must be closely integrated, and at all times care must be taken that elaborate and expensive space-vehicles are not used to obtain results that could be derived adequately by the use either of entirely ground-based techniques or of balloons. Equally, it is necessary for scientists not directly concerned in space research to be thoroughly aware both of the contributions already made to their subject by it and of the possibilities it affords for further investigations. To realize the potentialities fully it is necessary to set up laboratory programmes directed on the one hand towards the development of appropriate new techniques of measurement, and on the other to supplementing and interpreting information from space research.

TYPES OF SPACE-RESEARCH VEHICLE

We have already implicitly and arbitrarily indicated that we do not regard a balloon as a space vehicle. Conforming to the idea of a unity of technique, we regard as a space-research vehicle one that requires, at some stage or other, rocket propulsion independent of the existence of a surrounding atmosphere. Within this category it is best to distinguish between vehicles that pursue nearly vertical trajectories, returning to earth within a few minutes of launching, and those that leave the Earth's surface for long periods, and even for ever, as artificial satellites or deep-space probes.

All these classes of vehicle have an important role to play. The vertical sounding rocket has the advantage of relative cheapness, but suffers from the shortness of the period during which observations may be made. Many of the properties and phenomena of the upper atmosphere vary greatly according to place and time, and it therefore requires an enormous number of vertical probing flights to obtain a thorough knowledge of any of them. Further, there are weak effects that cannot be detected in a short flight and can be studied only by long-term observation. The use of artificial satellites, when possible, obviates these difficulties. Thus a satellite circulating for a month can provide as much data as eight thousand vertical sounding rockets using similar equipment: there is nearly world-wide coverage, and both sporadic and weak effects can be studied effectively. On the other hand, a satellite cannot circulate for long at altitudes much below 150 km because of air drag; this quite soon brings the orbit within the dense lower atmosphere, where the satellite burns up.

It follows that vertical sounding rockets are essential for studies of the atmosphere between the upper limit attainable by balloons (about 30 km) and 150 km. To overcome the chief disadvantage of such rockets, attention must be devoted to simplifying their design and launching, so that they may be fired frequently from many locations, including ones over the sea. Full advantage can be taken of such developments only if there is a corresponding reduction in the size and weight of the scientific equipment necessary to make any particular kind of observation.

For studies at altitudes above 150 km and requiring systematic observations over considerable periods—and this applies to most investigations—artificial satellites are the appropriate vehicles. There are some phenomena, however, that may be adequately studied at altitudes above 150 km even in the short time of a rocket flight; certain cosmic-ray and radioastronomical measurements are of this kind. In such cases it is not economical to use satellites.

For most purposes it is not necessary to recover either the instruments or records of any kind, the data obtained in flight being transmitted back to receiving stations on the ground in the form of coded radio signals. But recovery is desirable in those cases in which it is an advantage to use photographic film. There is no serious problem involved in doing this after a vertical sounding flight, but it is much more difficult to bring back equipment within a satellite. However, as this is a problem that must be solved before man can venture forth in a satellite, the Soviet Union and the United States have studied it closely, and both have achieved successful recoveries.

To explore the region between the Earth and the Moon (cislunar space) and to study the Moon and planets, it is necessary to launch vehicles at speeds close to the escape value (25 000 miles per hour). Such vehicles may be arranged to pass round the Moon and return to the near neighbourhood of the Earth; to hit the Moon; to pass close to the Moon and then continue on to become an artificial planet; to become an artificial lunar satellite; or merely to pursue a highly eccentric elliptic orbit round the Earth so as to penetrate at the apogee to distances of several earth radii into cislunar space. The possibilities may be enlarged considerably if one includes 'soft' lunar landings and planetary journeys similar generally to those we have listed for the Moon. One specially interesting and important case is that of a satellite stationary with respect to points on the Earth's surface. This will be the case for a satellite revolving in a nearly circular orbit in the equatorial plane with a period exactly equal to that of the Earth's rotation on its axis, a condition satisfied if the satellite is at a distance of about 23 000 miles from the Earth's surface.

TECHNICAL REQUIREMENTS

Apart from the obvious need for sufficiently powerful rocket motors and accurate systems of guidance and control, there are many other technical requirements. Some of these are of a general character common to all space research, while others are peculiar to the actual experiments that determine the payload of the vehicle concerned.

We have already mentioned the necessity of providing means for accurate tracking after launching. The most precise means of doing this is to use optical methods when these are applicable, but in many circumstances it is necessary to rely on radio direction-finding based on a transmitter within the vehicle. An allied problem is that of telemetry, that is, the transmission of data by means of coded radio signals from the vehicle. It is clearly necessary to extend as far as possible the range both of useful radio direction-finding and of data transmission. To assist in this, the necessary equipment should be both sensitive and capable of high discrimination between the wanted signal and background noise. The recent trial carried out jointly by the National Aeronautics and Space Agency of the United States and the Jodrell Bank Experimental Station has shown what can now be achieved. Not only were signals received from a space vehicle up to a distance of 22 000 000 miles, but the transmitter in the vehicle was switched on by command signal from Jodrell Bank when at a distance of 8 000 000 miles.

Another vital technique that must be developed is that of controlling the temperature of the vehicle so that it remains within the range in which the instruments on board will operate adequately. This problem is not very serious with vertical sounding rockets, but is important for satellites. The temperature at any time depends essentially on the balance between emission and absorption of radiation. Not only must it be controlled within prescribed limits but it must not fluctuate too much as the satellite moves in and out of sunlight. Much progress has been made in dealing with this problem by coating satellites with material of appropriate emissive properties. The problem is more severe for a vehicle penetrating to the neighbourhood of Venus, as the intensity of sunlight appreciably increases in this closer approach to the Sun. As indicated earlier, recovery of a satellite is difficult, and one of the main difficulties is the severe heating resulting from atmospheric resistance; here the temperature-control problem is much more serious.

Power must be supplied within a space vehicle to operate the measuring instruments and the radio transmitters. Here again there is no difficulty in a vertical sounding rocket, as conventional batteries may be used; but for satellites designed to circulate for a year the necessary battery capacity can be obtained only at the cost of a prohibitive increase in total payload. One effective solution is to use cells activated by sunlight. Another suggestion is to make use of radioactive sources, and there is clearly a great incentive to extensive research in this direction.

Little space research could be done if it were not for the availability of high-speed methods of computation. These make it possible rapidly to predict from the early observations the future track of a fast-moving vehicle, and this is essential for guidance and control systems. An equally important application is in the analysis of the vast amount of data transmitted from equipment in a satellite that has been circulating for months. The importance of this cannot be exaggerated, for unless the desired information can be extracted rapidly from the original records the whole operation is made valueless.

Many other technical problems arise in considering in detail the contents of a space vehicle that is to perform selected observations. There is, of course, the requirement of high reliability of the equipment, without maintenance, and there is the need for keeping to a minimum the volume and weight of the different components. Countless other problems have also to be resolved. Adequate methods of laboratory testing of prototypes under simulated conditions of acceleration, temperature, and so on must be developed. At all stages there is great scope and need for technical ingenuity, and the solution of the widely varying, unusual, and difficult problems associated with space research is bound to lead to new techniques of value in many other directions. To demonstrate this one need only point to the question of reliability of automatic operation over long periods.

Proceeding now to discuss the nature of the possible applications to different aspects of pure scientific research, it is perhaps best to begin with the prospects space research affords for advancing the study of the Earth's atmosphere and then to move gradually outwards. This approach is not intended to suggest an order of importance.

THE EARTH'S UPPER ATMOSPHERE

Above the limit of about 30 km attainable by balloons the atmosphere extends for hundreds of kilometres before merging with interplanetary space. It is best to judge the extent to which the atmosphere is still present at any altitude not by its pressure but by the number of atoms, ions, or electrons present per unit volume. Although at 100 km altitude the air pressure is about one-millionth of that at ground level, there are still about 5×10^{13} atoms and molecules per cubic centimetre. At 300 km, where the air pressure has fallen to 10^{-11} times the ground value, there are still more than a million times as many atoms and molecules than there are neutral or ionized particles in the same volume of cislunar space.

The atmosphere at great heights is the site of many phenomena, some of which are of considerable importance for us on the ground. The quickest way of summarizing these phenomena is by relating them to the solar source—the Sun exerts a dominant influence on the atmosphere through the radiations it emits. The Moon, too, influences some phenomena by producing lunar atmospheric tides. As the Sun is a variable star, the nature and intensity of its radiation are not steady, but we may distinguish a general background, characteristic of the Sun when quiet, from the additional radiations emitted under disturbed conditions. Such disturbances are usually associated with variations in number and size of sunspots and are most frequent at intervals of eleven years. In fact, the solar activity, measured in terms of numbers of sunspots, goes through an eleven-year cycle.

The radiation characteristic of the quiet sun is electromagnetic, and its intensity in the visible region is what would be expected if the surface temperature were about 6000° C. Effectively, only the visible radiation reaches the ground; other wavelengths, both shorter and longer (except for narrow bands in the infra-red and radio waves between 1 cm and 15 m wavelength (see figure 1)), are absorbed at high altitudes and in consequence produce important atmospheric effects. These effects include the formation and maintenance of the ionosphere, which extends upwards from an altitude of 100 km to several times that height. The ionosphere is a region in which there is a considerable concentration of free electrons (10^5–10^6 per cubic centimetre), sufficient to affect markedly the propagation of radio waves longer than about 15 m, so that instead of passing out along straight paths into space they are reflected back to the ground, thus rendering possible long-distance radio communication. The supply of free electrons is maintained through ionization of atmospheric atoms and molecules by solar ultraviolet light and X-rays.

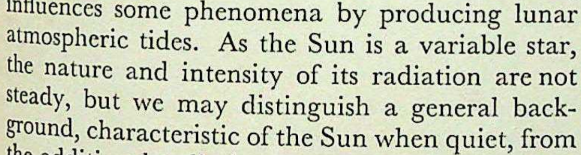


FIGURE 1 – *Spectrum of electromagnetic waves, showing the wavelengths (unshaded) that can penetrate the entire thickness of the Earth's atmosphere.*

Another remarkable effect of the steady ultraviolet radiation from the Sun is on the composition of the atmosphere. In consequence of photodissociation by these rays, oxygen exists predominantly in the atomic form at altitudes above 100 km. At lower altitudes, not far from the balloon limit, solar ultra-violet light produces photochemical effects that lead to a relatively high concentration of ozone. The ozone layer absorbs the remaining ultra-violet rays, which would otherwise be lethal for plant life on the ground.

These are just two of the major effects of solar radiation. There are many others, including the generation of the faint nocturnal airglow[1] and the quiet magnetic variations that arise from solar and lunar atmospheric tidal motions in the ionosphere.

The most conspicuous feature of the disturbed Sun is the emission of streams of charged particles, mainly protons and electrons, with speeds of about 1500 km/sec but at times including more energetic particles. Because of their charge, these particles are influenced by the Earth's magnetic field, so that they tend to enter the atmosphere in narrow belts located about 23° from either geomagnetic pole. In these regions they are responsible for the magnificent auroral displays, as well as strong disturbances of the Earth's magnetic field and of the ionosphere that may extend over a much wider latitude range. Most of the solar particles are not energetic enough to penetrate to within 80–100 km of the ground.

---

[1] The atmosphere at night emits a faint glow, known as the airglow, that is independent of auroral disturbances and is associated with photochemical reactions that arise from the effect of the Sun during the day.

Although we have merely indicated a few of the major aspects of the high atmosphere and of the nature of the solar influence, it should be apparent that there is a great deal of interest and importance to study. It is true that much can be done with entirely ground-based equipment, but if we depended on this alone many of the vital links in the chain of atmospheric processes would always remain obscure. Foremost among these is the solar radiation, which produces the atmospheric effects and, as it is absorbed in so doing, never penetrates into the lower atmosphere at all and cannot therefore be studied at ground level. We must have information about its nature and intensity before we can begin to understand the mechanism of production of the ionosphere, the airglow, the aurora, magnetic storms, and so on.

There are many other properties and phenomena that cannot be observed directly except with instruments taken to high altitudes. We need only cite the distribution of the positively charged particles in the ionosphere; the proportion of oxygen in atomic form at different altitudes; and the extent to which the atmospheric components separate out at different altitudes because of diffusion, to realize what a wealth of new possibilities arise, particularly when it is remembered that atmospheric properties vary with position over the surface, and with time of day, season, and position in the sunspot cycle. Furthermore, although information about atmospheric pressure, density, temperature, wind distribution, and electron concentration as a function of height can be derived—at least over certain altitude ranges—from observations on the ground, much more detailed and extensive information on all these variables can be obtained by direct measurement from a space vehicle.

The motion of a stream of charged particles from the Sun under the influence of the Earth's magnetism is so complicated as to defy theoretical prediction; but it is now possible, through the use of particle detectors and magnetometers carried in deep-space probes, to extend the study of the behaviour of these streams to great distances from the Earth. Perhaps the most important discovery yet made in space research has been that of the radiation belts surrounding the Earth (see figure 2). These are regions in which there is an unexpectedly high density of moving charged particles having energies of the order of a few tens of thousands of electron-volts. In some way the production and presence of these belts are associated with auroral and magnetic-storm phenomena.

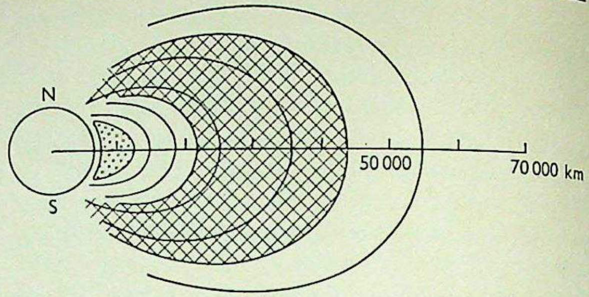


FIGURE 2 – *Natural radiation belts surrounding the Earth. The inner belt is indicated by stippling and the outer by cross-hatching; lines of force of the Earth's magnetic field are also shown. The outer belt is variable, and the region shown is based on observations made with equipment in Luniks I and II.*

Although there is no fully adequate understanding of how the supply of particles is maintained, nor of how the particle concentration varies with solar activity, it seems fairly certain that the effect of the Earth's magnetic field in trapping particles moving with certain velocities is of major importance. The thorough study of the radiation belts is likely greatly to improve our knowledge of the way in which the Sun affects atmospheric phenomena. It can be carried out only through space research, and will take many years of systematic observations.

PROPERTIES OF THE SOLID EARTH

The orbit of an artificial satellite is sensitive to the shape and constitution of the solid Earth, because this determines in detail the gravitational field acting on a body at any point. Already new data relevant to this have been obtained from analyses of the orbits of artificial satellites and much more will be learned as tracking methods become more precise and continuous. In this connection it will be possible to improve the accuracy of geodetic surveys—a matter of practical as well as scientific importance.

APPLICATIONS IN GENERAL ASTRONOMY

It has been pointed out above that only visible sunlight and certain solar radio-waves are able to penetrate the atmosphere, and this is necessarily true also of the radiation from the stars. Virtually all astronomical studies depend on observation of electromagnetic waves from heavenly objects, covering less than two octaves in the visible and about the same range in the radio region (figure 1). Observations made from outside the atmosphere are not so limited. They may be extended by many octaves to higher frequencies, through

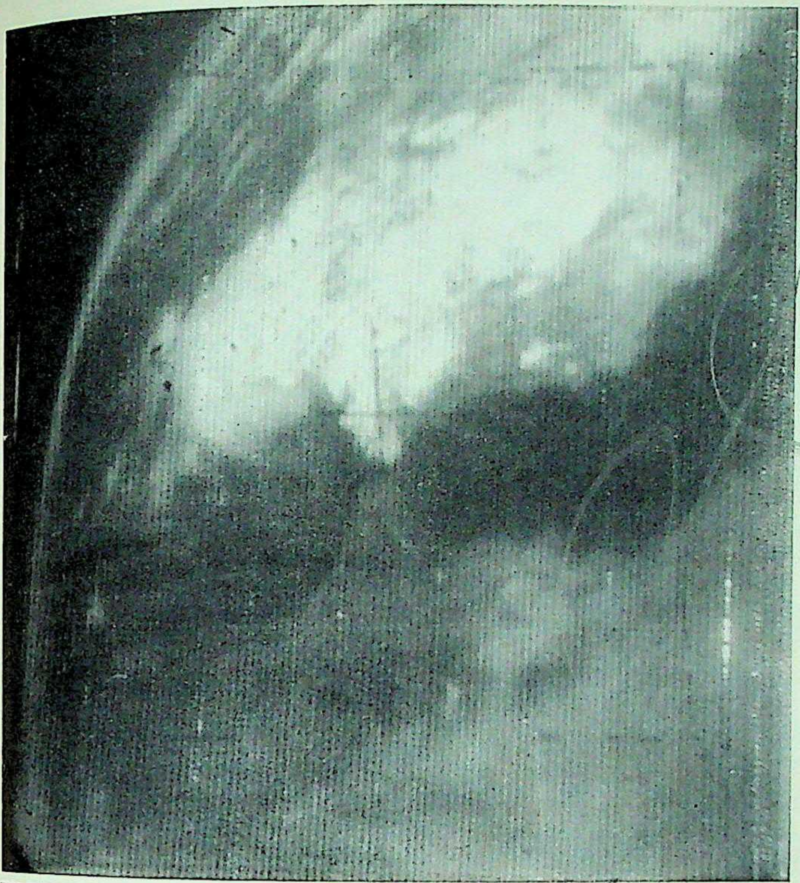
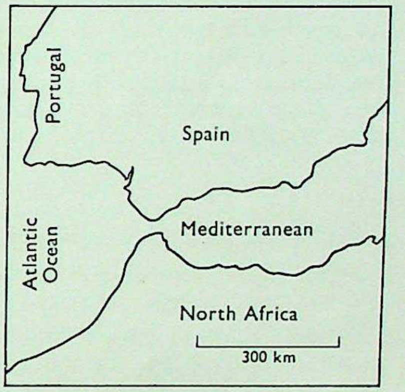
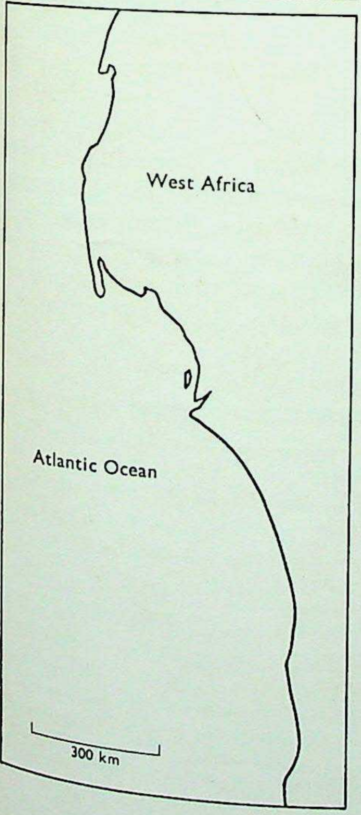


FIGURE 3 – *Photograph taken from Tiros I (launched 1st April 1960) while over North Africa, camera pointing downwards and to the west. The Straits of Gibraltar can be seen in the centre of the picture, with the Mediterranean to the right and the Atlantic to the left. Note cloud over Portugal.*


Portugal
Spain
Atlantic Ocean
Mediterranean
North Africa
300 km


West Africa
Atlantic Ocean
300 km

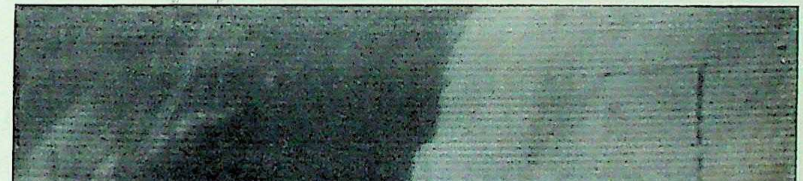


FIGURE 4 – *Part of the west African coast, approximately from Cape Blanco to St. Louis, photographed from Tiros I.*

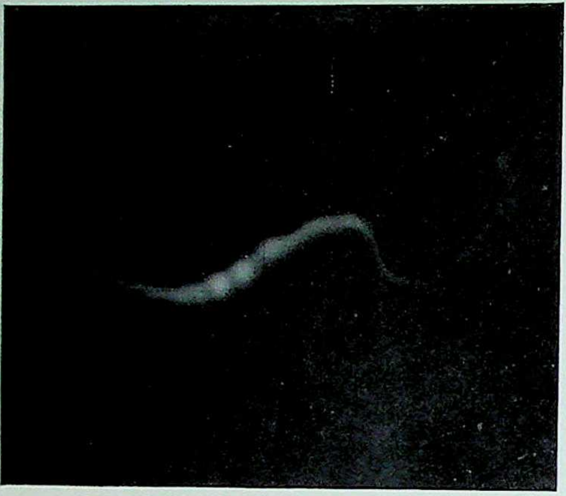


FIGURE 5 – *Photograph of the fluorescing trail of sodium vapour ejected from a Skylark rocket launched at twilight from Woomera, Australia. By studying the variation in the form of the trail with time, information may be obtained about the wind distribution at the height of the trail (between 93 and 120 km).*

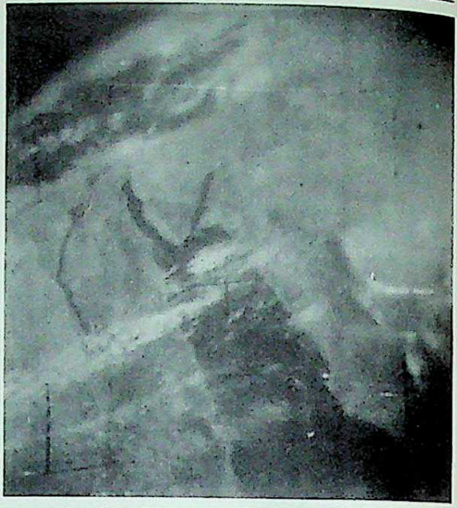


FIGURE 6 – *Photograph from Tiros I taken over the Red Sea, showing the Gulf of Suez and the Gulf of Aqaba, and eastern Mediterranean (top left). Cloud formations can be clearly seen across centre of picture.*

the ultra-violet and X-ray region to γ-rays, and equally to lower frequencies through the infra-red to long radio-waves (see figure 7). It is not possible to predict what will be discovered by means of telescopes or equivalent equipment, mounted in satellites, that view the universe in these hitherto inaccessible wavelength regions. Already preliminary scans of the northern sky have been carried out in ultra-violet radiation of wavelength 1500 Å (the shortest wavelength that can reach the ground is 2900 Å). Instruments in a vertical sounding rocket were used, and it was found that some of the regions of the sky that appear bright when viewed in this ultra-violet light do not appear so when viewed from the ground in visible light; this is a portent of new discoveries to come. It is important to note that the exploitation of these possibilities does not require the use of deep-space probes, as a satellite circulating a few hundred miles up will be quite adequately outside the atmosphere (see figure 7). More elaborate equipment, such as an instrument platform automatically stabilized in a plane effectively fixed with respect to the stars, will be an essential requirement as the subject develops; but this does not represent too serious a technical problem.

METEOROLOGICAL APPLICATIONS

So far we have been concerned with observations looking outwards from the Earth, but much of value can be obtained by looking inwards at the lower atmosphere with equipment carried in an artificial satellite. In this way the cloud cover over the whole Earth can be under continuous observation and features undetected from the ground may be clearly discerned (figures 3, 4, 6). These include the development of tropical storms and other disturbances; information of this kind cannot fail to assist meteorologists in day to day weather prediction. Moreover, in association with regular observations of the gain and loss of heat from the atmosphere, which can also be made with satellite-borne equipment, it may even lead to much deeper knowledge and understanding of atmospheric circulation, so making long-range weather forecasting possible.

LUNAR PHYSICS

The Moon is an especially interesting object to

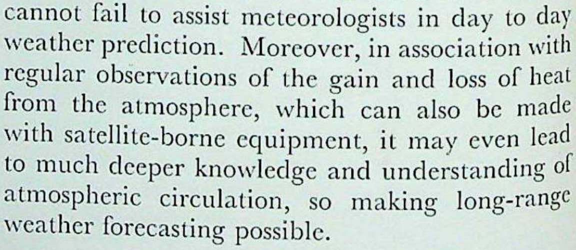


FIGURE 7 – *Height in the atmosphere at which ultra-violet- and X-radiation entering from outside is reduced in intensity to 1/2·7 of its initial value. The corresponding height for X-rays is indicated by* **x**.

study because of the light that may be thrown on the origin of the solar system. The thermal history of the Moon—whether it has ever been molten—is of particular interest. In this context, the measurement of lunar radioactivity and magnetism, if any; the determination of the chemical composition of surface material; verification of the presence or otherwise of volatile constituents; and detailed investigation of the nature of the surface are all likely to prove of much value, besides being of interest in themselves. Thus selenology would be expected to involve quite different considerations from those of geology because, apart from anything else, there has been neither erosion by water nor production of sedimentary rocks on the Moon. Moreover the exposure of the surface to meteor bombardment and short-wave radiations, as well as to the effects of extreme changes of temperature, will have introduced features unknown on the Earth.

In the first instance, lunar observations will be made from vehicles that do not land—a task begun with the Russian Luniks and the American Pioneer probes. Already there is evidence that the lunar magnetic field is very weak; but so far, except for the Russian lunar impact, the probes have not approached closer than a few thousand miles from the Moon. Before long a vehicle will be placed in a near orbit, in which it will make a considerable number of revolutions round the Moon.

From accurate observation of the orbits of such lunar satellites, new information about the shape and composition of the Moon will be forthcoming. If the orbit is sufficiently close (within about 100 km) it will be possible to investigate the spectrum of the γ-rays emitted from the Moon. This will indicate the composition of the radioactive sources responsible, which will in turn give information about the nature of the surface rocks. Tests for the presence of a thin surface-atmosphere or ionosphere, and magnetic measurements, could also be carried out from a close lunar satellite. Echo-sounding would be employed to give further information about the nature of the surface—a matter important for future landing of equipment. Television pictures of the surface would naturally play a vital part throughout.

The next stage would involve a fairly hard, but not too hard, landing of equipment. It should be possible to design a robust seismograph which would survive the initial shock and transmit back to Earth data about the occurrence and scale of moonquakes, if any. With development of technique, so that landing shocks were greatly reduced, more elaborate apparatus could be set in operation. For example, an X-ray analysis of the surface could be carried out by bombarding it with an electron beam and observing the fluorescent X-ray spectrum emitted. This is a mere foretaste of possibilities, the scope of which is only gradually being appreciated through the increased attention being paid to the subject.

PLANETARY STUDIES

Many of the remarks made about the Moon apply to the planets, of which only Mars and Venus are likely to be in range for many years to come. However, there is an important difference from the Moon, in that both possess extensive atmospheres that would greatly complicate the task of landing equipment safely on the surface: heating due to friction of the fast-moving vehicle with the atmosphere would be severe. For a considerable time attention would therefore have to be concentrated on studies of the atmosphere and ionosphere and of surface conditions, with probes circulating a few hundred kilometres from the surface.

Because of the biological information that might be forthcoming, especial importance attaches to the soft landing of equipment. The question of whether there exists, in an extra-terrestrial environment, anything that could be called living is of supreme interest in relation to the wider problem of the nature and origin of life. There is evidence suggesting the presence of some form of lower plant life on Mars, so that there are real prospects of major discoveries when information can be received about the nature of the surrounding material from instruments placed on the Martian surface. Venus may turn out to be even more surprising, as we know little about the surface, or even about the atmosphere, of this planet.

OTHER SCIENTIFIC APPLICATIONS

Although the variety and scope of the scientific work already outlined is very great, it is by no means comprehensive. There are other less obvious possibilities, such as the study of the nature of gravitation. This force is the weakest of the four major types of natural force known to us and is apparent in everyday life only because of the ubiquitous presence of a very large body, the Earth; the other three forces can be observed even between fundamental particles with masses of 10^{-24} g or less. Because it is so weak we know less about the nature of gravitation than about the other three types of force; not that our knowledge of these is very profound. With space probes available there will be opportunities for precise observations, under controlled conditions, of small orbital perturbations and of the relative rates of clocks in different gravitational environments. The prospects in these directions, though vague at present, may well turn out to be of basic scientific importance.

The most important new results that will be obtained through space research will be unexpected ones. All that has been done here by way of illustration is to list the broad headings within which investigations will proceed and some of the ways in which they will begin. As the work expands, more and more ingenious proposals will be made and implemented.

ATMOSPHERIC AND COSMIC EXPERIMENTS

So far we have been describing what types of scientific observation may be made. Hitherto all upper atmospheric and astronomical research has had to be of an observational kind, and much future work, including that carried out from space vehicles, will continue to be so. On the other hand, it is now possible to carry out experimental investigations in which conditions are deliberately modified in some controlled way and the consequences studied.

Thus the upper atmosphere may be used as a vast photochemical laboratory, free from surface effects. Various chosen materials may be ejected into the surrounding air at suitable heights, producing optical and electrical effects that can be observed from the ground (see figure 5). Thus a few pounds of sodium vapour ejected at twilight at a height of about 100 km produces a yellow fluorescent glow that lasts up to thirty minutes; a similar weight of nitric oxide introduced at about 120 km reacts, photochemically, quite vigorously with the atomic oxygen; and so on. In daytime, artificial ionized clouds may be similarly produced. Already a good beginning has been made in carrying out this kind of atmospheric experiment.

By far the most remarkable large-scale experiment was successfully carried out by the United States in 1958, the 'Argus' experiment. On three separate occasions, belts of particle radiation having an intensity comparable with that of the natural belts were produced artificially by exploding atom bombs of conventional type at an approximate altitude of 480 km above a point in the South Atlantic. The charged particles produced in these explosions spread round the Earth in a belt 40–100 km thick. These belts lasted for some days, during which they were studied with equipment aboard an American satellite (see figure 8). Besides the radiation belts, artificial auroral displays were observed both above the firing site and at the conjugate[1] magnetic location in the northern hemisphere (the Azores). This was an experiment on the cosmic scale and plainly indicates what immense possibilities are now opening before us.

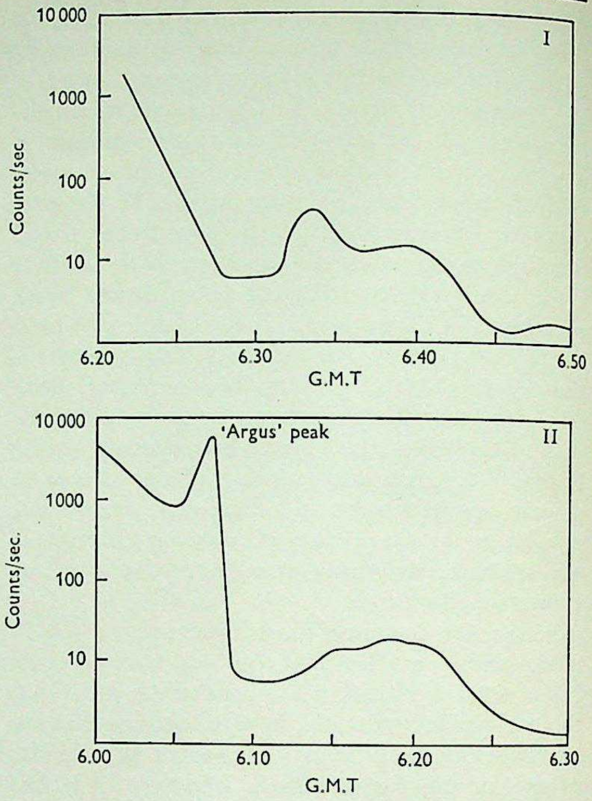


FIGURE 8 – *Results of observation of the radiation belt produced artificially in the first 'Argus' experiment. I and II show the variation with time of the counting rate of the unshielded counter in the Explorer IV satellite. I refers to rates found on 26th August 1958, the day before the experiment; II to the corresponding rates the next day, about 3½ hours after the explosion. The peak due to the particles released from this explosion is indicated.*

COMMERCIAL APPLICATIONS

The last remarks apply with equal force to any commercial applications. It is too early to predict most of these, although the practicability of large-scale modification of the Earth's environment has already been demonstrated by the Argus experiments. Furthermore, it is already clear that there are important applications of satellites to world-wide communications.

---

[1] The conjugate point is that at which the line of magnetic force through the firing point returns to the atmosphere in the other hemisphere.

At present long-distance communication by radio depends on the ability of the ionosphere to prevent the radio waves from passing tangentially out into space, but the ionosphere is effective in this way only for waves that are longer than 15 m. Shorter waves, such as are used in television, cannot, therefore, be used for long-distance transmission. Apart from this there are other limitations: the ionosphere is highly variable, particularly at high latitudes, so that communication is capricious. For these reasons it would be advantageous to replace the ionosphere by something more reliable and less dependent on the use of a particular wavelength.

One way of doing this would be to set up a number of satellites that would act as reflectors of radio waves. Such a passive system would have attendant disadvantages, many of which could be overcome by using active satellites that would receive signals and retransmit them back to the Earth. These satellites would be in such orbits that they remained fixed with respect to points on the ground. It would be necessary to include devices that could be operated on command to restore them to their correct positions from time to time as they drifted away as a result of various perturbations. However, it seems possible that a network of this kind would be a good commercial proposition, even allowing for regular replacement launchings. A further possibility is to lay down an artificial ionosphere of small conducting needles that would form a belt round the Earth at an appropriate altitude. By choosing the size of the needles this belt could form a reflecting layer for short radio-waves in a narrow wavelength band; the total weight of material required would not be prohibitive. This suggestion is, however, looked on with horror by radioastronomers, who fear that their opportunities to observe the relatively very weak signals from extra-terrestrial sources will be restricted even more by disturbances due to artificial reflecting belts of this type. But all this serves to show what big, and sometimes dangerous, possibilities can now arise.

CONCLUSION

Space research covers an extremely wide field of scientific endeavour, and its importance in the context of pure science is very great. For its success, it is necessary to develop and improve continually a wide range of techniques, including rocket propulsion, power sources, communication and other forms of electronic engineering, high-speed computation, and automation generally. The instruments required must be especially small in volume and weight, robust, and capable of operating reliably over long periods without maintenance. It is obvious that the by-products of these varied technical developments will be of much importance and value in many other fields. Apart from this, there are already forseeable practical applications, as in the field of communications, and there will undoubtedly be many more as the work develops.

Expenditure on space research is therefore likely to be a sound investment provided it is maintained within appropriate limits. This kind of research is expensive and potentially unlimited, but it is not necessary to embark on all types of activity in the field in order to obtain some dividends; for any individual country the problem is to decide how far it should go. This is obviously very difficult in any particular case, and it is clearly desirable that the maximum amount of international co-operation should be encouraged. Because of the great expense involved in providing the basic tools, space research must be highly organized; but it is essential that the organization should not be so cramping that there would be no opportunity for individual enterprise, so important in all scientific research.

ACKNOWLEDGMENT

I am particularly grateful to Dr Frutkin of the National Aeronautic and Space Agency, U.S.A., for providing the photographs taken from Tiros I.

Fluorescent protein tracing and the fluorescent antibody method

R. C. NAIRN

Fluorescent dyes may be used for labelling proteins without materially affecting their biological or immunological properties. This is most valuable for biological research, for such labelled proteins can be injected into animals and traced directly in histological sections by ultraviolet fluorescence microscopy. Alternatively, in immunological tracing, labelled serum-antibody is used as a specific histochemical stain to locate the corresponding antigen in microscopical preparations: this article describes and illustrates the study in this way of a wide variety of antigens, including micro-organisms and protein constituents of tissues.

There can be no doubt about the value to biology of a method that permits the tracing of native or foreign proteins in the tissues of living organisms. Some proteins can be labelled and traced directly, others by the application of immunological principles in which they are located histologically by means of labelled specific antibody. Such labels are provided by certain fluorescent dyes which may be combined chemically with proteins, including serum antibodies, without material effect on the biological or immunological properties of the proteins. Fluorescein [4] and lissamine rhodamine B (RB200) [2] are the fluorochromes in commonest use at present, giving respectively conjugates with apple-green and orange fluorescence in ultraviolet light. In suitable preparations of tissues, cells, and micro-organisms, these conjugates can be made visible by ultraviolet fluorescence microscopy, with considerable histological precision.

DIRECT TRACING

The fluorescent conjugates of serum proteins, when injected into animals, seem in all important respects to be accepted biologically as normal proteins [26]. They are not toxic, and their distribution in the body and eventual elimination are much the same as with proteins labelled with radioisotopes. The distribution of the proteins can be demonstrated by fluorescence microscopy in histological preparations of tissues removed from the organism, or sometimes *in vivo* as in microcirculatory studies of mesenteric loops. Fluorescent tracing is a complementary technique to radioactive tracing: the sensitivity is substantially less but this is fully compensated by its greater convenience, speed, and histological precision.

IMMUNOLOGICAL TRACING

This is an immuno-histochemical technique quite distinct from direct tracing and much more widely used. It depends on the fact that serum antibody that has been labelled with fluorescein or RB200 still retains much of its immunological activity, usually about 50 per cent. Such fluorescent antibody can be used as a specific immunological stain for micro-organisms, proteins, and other macromolecules, which can therefore be identified even in the presence of closely related organisms or substances. The principles of the method are illustrated by an experiment in which an antigen, such as a suspension of micro-organisms, is injected into a rabbit to stimulate antibody production. The antibody, formed after two or three injections over a few weeks, is present in the γ-globulin fraction of the serum which is conjugated with the fluorochrome. The conjugate and the corresponding antigen react with immunological specificity; the organisms, coated with fluorescent antibody, fluoresce brilliantly when illuminated by ultraviolet or ultraviolet-blue light. The fluorescence microscope used for this work is so designed that the light from the ultraviolet source is filtered off from the observer's eye, usually by means of a filter built into the ocular, and only the fluorescence of the specimen under examination is seen. Consequently, the apple-green fluorescence of fluorescein conjugates or the orange of RB200 conjugates is seen against a black background.

In theory, the method could be employed for visually identifying any substance capable of stimulating antibody production. In practice, two difficulties may arise. Firstly, the antigen may fail to provoke an adequate antibody response, or the presence of impurities in the antigen may give

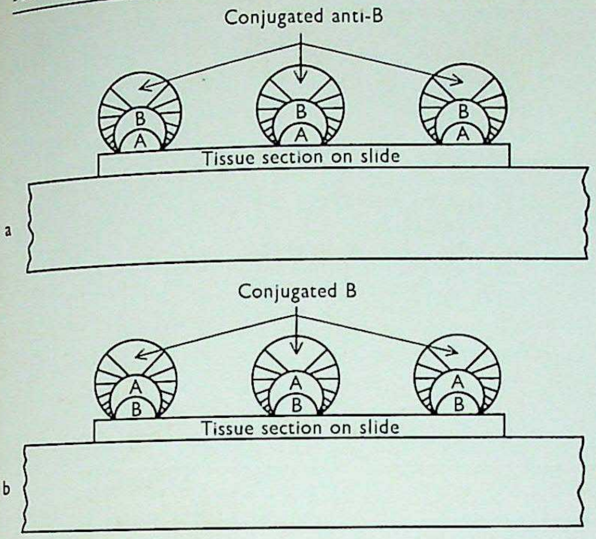


FIGURE I – *Double-layer (sandwich) methods in fluorescent antibody tracing.* (a) *Location of antigen (A) with layer of unconjugated specific antiserum (B) followed by a layer of conjugated antiglobulin. If, for example, B is rabbit antiserum, then the second layer would be anti-rabbit-globulin prepared in some other animal.* (b) *Location of antibody (B) by applying a layer of specific antigen (A) which is then made evident by a layer of conjugated B antiserum.*

rise to unwanted antibodies. Secondly, there may be inability to fix and preserve the antigen in preparations suitable for microscopy. These difficulties can often be overcome, and there is a growing literature about the applications and technical development of the fluorescent antibody method. The antigens that have been studied include tissue components such as connective-tissue elements, cytoplasmic and nuclear proteins, enzymes, and hormones, and also helminth parasites, protozoa, fungi, bacteria, and viruses. The localization of antigen is carried out in fresh-frozen or freeze-dried tissue sections, in smears, or in tissue culture preparations. For insoluble antigens such as tissue reticulin or collagen, preparations may be treated directly with the fluorescent antibody; soluble antigens require initial fixation with some histological fixative such as acetone or 95 per cent ethanol, which rarely destroys antigenicity. Conventional fixatives and dehydrating methods are avoided because they usually denature the antigens.

An alternative way of applying the antibody technique is by the double layer or 'sandwich' method (figure 1*a*). The antigen (A) is first coated with unlabelled specific antibody globulin (B) which in turn is located with a labelled antiglobulin serum. This method has the advantage of greater sensitivity than the single-layer method and it reduces the number of conjugated sera required for work with several different antigens. For example, a single antiserum, prepared against rabbit globulin in an animal such as a goat, can be conjugated and used to detect any rabbit antiserum attached to its corresponding antigen. A different kind of sandwich technique may be used to locate antibody in animal tissues (figure 1*b*). In this, the preparation is treated with the corresponding antigen (A) which becomes attached to the antibody (B) in the tissues and can then be located with labelled antibody.

TECHNIQUE

For direct tracing, the conjugated serum protein is used at the same protein concentration as serum and injected intravenously in a dose of about 8 ml/kg body weight. The tracer is easily identifiable histologically by fluorescence microscopy up to six hours after injection [21].

For immunological tracing, conjugated antiserum is used and the first step is commonly the preparation of the specific antibody. The antigen to stimulate antibody production should be as pure as possible in order to avoid the production of unwanted antibodies; it may be injected in saline solution into suitable animals, but frequently is first mixed with an adjuvant, such as Freund's. Adjuvants retard the absorption of the antigen and also promote, by non-specific inflammatory reaction, the production of antibody-forming cells in the immunized animal. Usually two or three injections of about 5–50 mg of antigen per kilogram body weight at monthly intervals give a satisfactory antibody response, which may be measured by conventional *in vitro* tests such as agglutination, precipitation, or complement fixation. The antiserum, which often contains as much as 3 mg of antibody globulin per millilitre, is conjugated with the fluorochrome; the conjugate is treated with ammonium sulphate (40 per cent saturation) to precipitate the γ-globulin and thus concentrate the fluorescent antibody, which is then resuspended in physiological saline and dialysed to remove the ammonium sulphate.

Chemical combination of serum protein with the fluorochrome is believed to take place largely through the free amino group of the lysine moieties. For this combination to take place, the fluorochrome molecule must contain a reactive group capable of forming a stable covalent linkage with the free amino-group. Fluorescein is used as the isocyanate or isothiocyanate [25], which form carbamido and thiocarbamido linkages respectively; RB200 is used as the sulphonyl chloride, *FIGURES 2–7 – These are fluorescence photomicrographs obtained by using ultra-violet-blue illumination with a yellow filter above the object; autofluorescence is greenish with this optical system.*

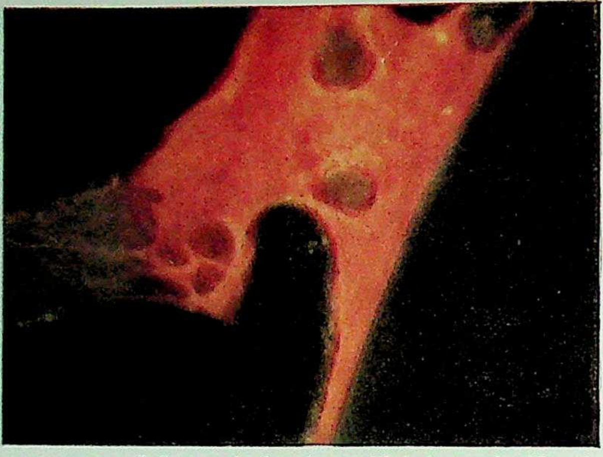


FIGURE 2 – *Influenza A virus in syncitial clump of calf testis cells stained with RB200 anti-influenza globulin. Specific fluorescence in cytoplasm but no staining of nuclei. Faint greenish fluorescence, left of infected clump, is an uninfected cell.* (×800)

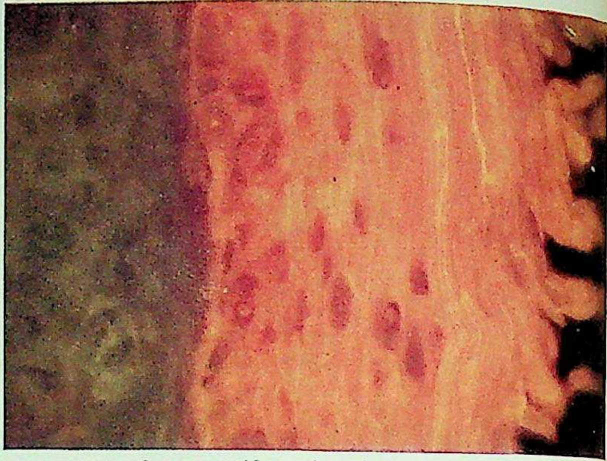


FIGURE 3 – *Organ-specific antigen in human epidermis to the right, with absence from underlying cancer to the left. Section of skin treated with anti-human-skin serum. Specific fluorescence is conspicuous in cell membranes.* (×550)

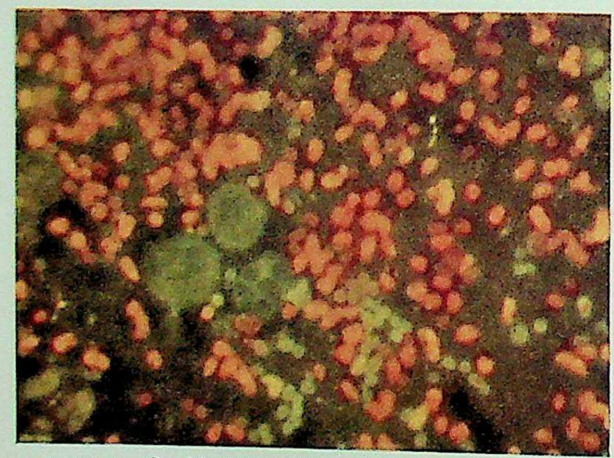


FIGURE 4 – *Combined klebsiella and pneumococcal infection in mouse. Peritoneal smear treated successively with RB200 anti-klebsiella globulin and fluorescein anti-pneumococcal globulin. Klebsiella stained orange and pneumococci green. Three greenish mesothelial cells left of centre.* (×800)

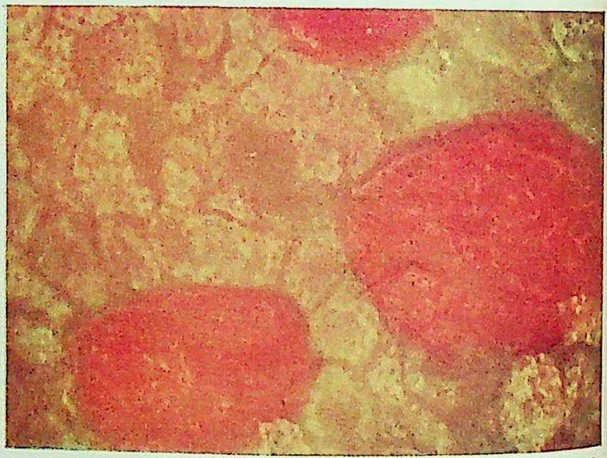


FIGURE 5 – *Renin in pig kidney. Section treated with anti-renin serum and then RB200 goat anti-rabbit globulin (sandwich method). Specific fluorescence in cells of the glomerula tuft and capsule; autofluorescence in the tubules.* (×180)

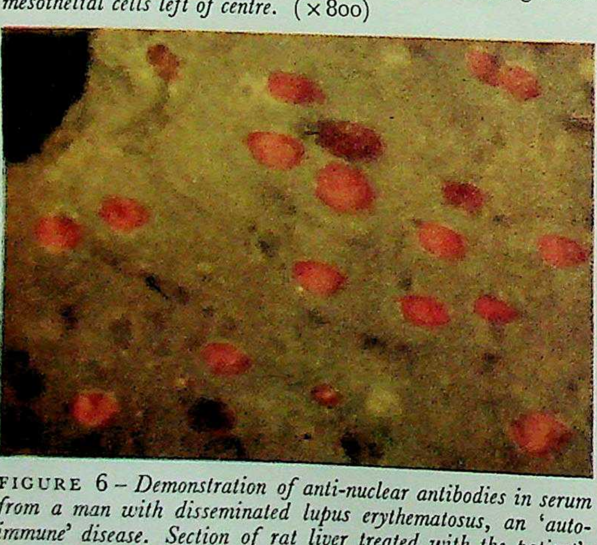


FIGURE 6 – *Demonstration of anti-nuclear antibodies in serum from a man with disseminated lupus erythematosus, an 'auto-immune' disease. Section of rat liver treated with the patient's serum conjugated with RB200. Specific nuclear fluorescence against greenish autofluorescent background; quadrantal space, top left, is a centrilobular vein.* (×550)

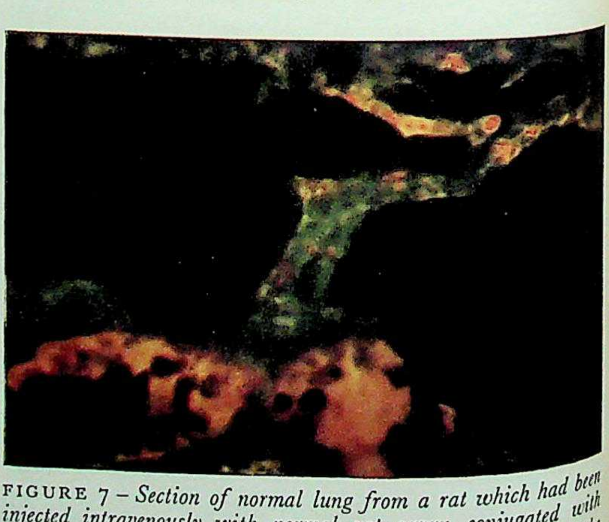


FIGURE 7 – *Section of normal lung from a rat which had been injected intravenously with normal rat serum conjugated with RB200. The orange fluorescent plasma contrasts well with the green autofluorescence of the lung tissue and the black non-fluorescent alveolar spaces; the small black bodies in the fluorescent plasma are red cells, which do not fluoresce.* (×550)

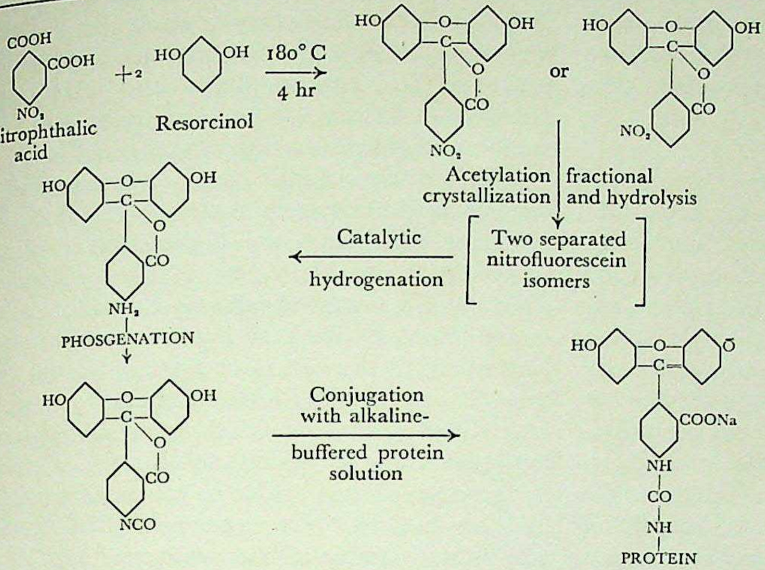

4-nitrophthalic acid
Resorcinol
180° C
4 hr
or
Acetylation crystallization
fractional and hydrolysis
Two separated nitrofluorescein isomers
Catalytic hydrogenation
PHOSGENATION
Conjugation with alkaline-buffered protein solution
PROTEIN

FIGURE 8 – *Preparation of fluorescein-conjugates. For isothiocyanate conjugation the more convenient liquid thiophosgene ($CSCl_2$) replaces the gaseous phosgene ($COCl_2$) and sulphur replaces the oxygen in the resulting reactive group.*

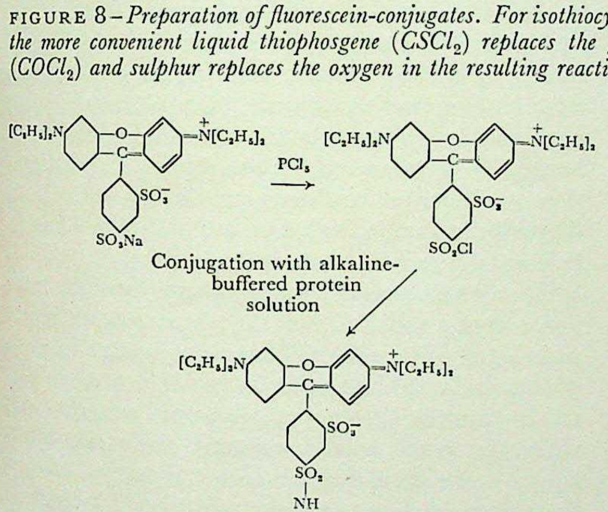


FIGURE 9 – *Preparation of conjugates with lissamine rhodamine B (RB200).*

which gives a sulphonamido linkage. The essential steps of the preparations are shown in figures 8 and 9. Fluorescein isothiocyanate is replacing isocyanate in practice because it is a stable substance and obtainable commercially, but it is expensive—about £10 per g compared with less than 6d. for RB200.

An excess of dye is used in the conjugation, and some of this is adsorbed by the serum proteins. Free dye in solution, and most of the adsorbed dye, may be removed by dialysis, by passage through resin columns, or by shaking with absorbing materials such as dried tissue powders; virtually all unconjugated dye can be removed by shaking with powdered activated charcoal [3]. The purpose of these absorption procedures is to obtain a fluorescent solution in which all the dye is firmly bound to the serum proteins. After centrifuging to remove particulate matter, conjugates are stored at – 20° C until required. For most purposes, antiserum conjugates need further absorptions with tissue powders or homogenates to remove unwanted antibodies or other proteins which may be a source of non-specific staining reactions. The conjugates are applied to the antigen, such as a tissue section, as a drop, and the preparation is kept in a damp container for 10–60 minutes to allow the reaction to take place. The surplus conjugate is then removed by washing in buffered saline, *p*H 7·0. A glycerol mountant, which is non-fluorescent, is used and the preparation is then ready for examination by fluorescence microscopy.

The microscopy demands a powerful source of ultraviolet-blue light. For this, the modern high-pressure mercury-vapour burner has largely replaced the carbon arc; it provides a small intense source with fairly constant output of wavelength 2750–6000 Å, with main peaks at 3650 Å and 4348 Å. Filters that pass ultraviolet or ultraviolet-blue transmit light up to 4100 Å or 4750 Å respectively, a suitable range for stimulating fluorescence in the fluorochrome labels in common use. Quartz lenses are not necessary for this long-wave ultraviolet light, but the collector lens and condenser of the microscope should be made of crown glass, which transmits light above 3000 Å, and surface-reflecting mirrors should be used. Dry objectives giving up to 60 × magnification are most convenient; when using oil-immersion objectives to obtain higher magnification, a dark-ground condenser and non-fluorescent oil are required.

The fluorescent antibody method permits precise histological localization of antigen at concentrations as low as 10 μg/ml [6]. Fluorescent tracers can be seen in much lower concentration (1 μg/ml) than is possible with non-fluorescent dyes. Although this is more than is required with radioactive tracers for autoradiography, the gain in histological precision with the fluorescent method is adequate compensation. Such precision often demands special equipment to obtain good microscopical preparations: for example, a low-temperature microtome cabinet is used to obtain uniform thin frozen sections.

RESULTS ACHIEVED

The direct tracing method has so far been little used. The physiological distribution of fluorescein-conjugated serum proteins was demonstrated by A. A. Schiller, R. W. Schayer, and E. L. Hess [26] in 1953, and an application of the technique to the study of experimental liver damage, using labelling with both fluorescein and RB200, was reported in 1958 [21]. J. Oort, in a recent personal communication, has described the use of human serum albumin labelled with RB200 as a tracer in the investigation of a rare disease of childhood characterized by protein-losing diarrhoea. Figure 7 demonstrates the kind of histological picture obtained by the direct method. Fluorescent plasma is shown in the capillaries of an unstained paraffin section of normal lung from a rat that had been injected intravenously one hour before with normal rat serum conjugated with RB200.

Immunological fluorescent tracing has had many applications in biological and medical research. Its value to microbiologists and immunologists has been unquestioned since A. H. Coons and his colleagues published their important series of studies on the distribution of foreign antigens, on the natural history of virus infections, and on antibody formation [5, 6]. The simplest example of immunological tracing is perhaps the specific staining of bacteria by a conjugated antibacterial serum. The term specific staining here denotes the process of coating antigen with fluorescent antibody and of proving by suitable tests that the reaction is specific. Control tests commonly used are: (1) that fluorescent staining occurs only when the corresponding antigen is present; (2) that no staining of the antigen occurs when conjugated normal serum is used; (3) that staining is inhibited if the antigen is pretreated with unconjugated specific antiserum; (4) that staining is inhibited if the conjugated antiserum is first absorbed (neutralized *in vitro*) with the corresponding antigen.

In medical research, there are already numerous examples of the method having been used to identify a specific organism in a mixed flora—for example, Group A streptococci in throat washings [19]. The method has also been used in veterinary research for the rapid identification of bacterial flora in the rumen contents of sheep and calves [11]. Similar methods may be applied to botanical research: specific soil organisms have been demonstrated in and around plants, and their spatial distribution defined, despite the presence of large numbers of contaminating organisms.

A refinement of this kind of immunological tracing is illustrated in figure 4, in which two different organisms are demonstrated simultaneously by double staining. The photomicrograph is of a smear of peritoneal exudate from a mouse infected with both a pneumococcus and a klebsiella. The smear, after acetone-fixation, was treated with anti-klebsiella globulin conjugated with RB200 and then with fluorescein-conjugated anti-pneumococcal globulin; the two bacterial antigens are clearly visible in contrasting colours. This method has also been used to distinguish two related strains of the protozoon *Trichomonas* in mixed smears [2]: cross absorption of the conjugates by the antigens is necessary in such experiments to obtain specific staining of each strain with its own conjugate only. Double tracing has obvious application to many immunological problems; it has been used for the separate identification of different antibody-producing cells in a single lymph node [29], and for the study of antibodies to closely related antigens [27].

The remaining photomicrographs illustrate some other applications of the fluorescent antibody technique. The use of RB200-conjugates in each of these reflects the author's special interest in this tracer; fluorescein-conjugates would serve as well wherever good colour contrast with the autofluorescence of the tissue is not essential. Autofluorescence is a natural phenomenon in biological material: it is usually blue-green in animal and plant tissue, except for the green parts of plants, which exhibit the red fluorescence of chlorophyll. Fluorescein does not provide good colour contrast with the normal blue-green autofluorescence of biological material, but this may not be important if the fluorescent intensity of the tracer is sufficient. When the intensity is poor, good colour contrast is essential and is provided by RB200.

Figure 2 demonstrates the value of the method for the localization of virus. It shows specific fluorescence due to the presence of influenza virus particles in the cytoplasm of an infected tissue-culture cell. The preparation is a coverslip monolayer fixed in acetone and stained with anti-influenza A globulin labelled with RB200. Such preparations offer a method of studying the natural history of virus infection—its location and development. The cytological localization can be sufficiently precise to permit a clear distinction to be made between antigen accumulation on the cell membrane, in the cytoplasm, in perinuclear membrane, and in the nucleus. Similar location of influenza virus has been made in experimental animal tissues [8] and in human material [15]. Many viruses have been successfully investigated in this way [16] and much information has been gathered on infectivity, growth cycles, and spread. Without the fluorescent antibody method, such information would have been obtained only with difficulty or not at all.

Figure 5 is an example of the use of the technique for the location of an antigenic substance native to the tissues studied—in this case the pressor enzyme, renin, in pig kidney [22]. For this study, renin, prepared in as pure a state as practicable by chemical fractionation of pig-kidney extract, was injected with adjuvant into rabbits to provoke the formation of anti-renin serum. This was absorbed with pig-tissue preparations to remove unwanted anti-pig and anti-kidney components and to leave the anti-renin. The absorbed serum was then applied to frozen sections of pig kidney after fixation in 95 per cent ethanol; the site of its reaction with the renin was located by subsequent application of anti-rabbit globulin conjugated with RB200. Specific fluorescence was obtained in the cells of the glomerular tufts, but not if the anti-renin serum had been previously neutralized by mixing with the measured amount of renin. The findings do not demonstrate conclusively that the material located is renin, because the renin solution used for the neutralization test was contaminated by related proteins; pure renin is not yet available. Other enzymes and some hormones have been located histologically with fluorescent antibody—for example, pancreatic enzymes [17, 20], insulin [13], and pituitary growth-hormone [14]. Successfully demonstrated native antigens also include plasma proteins [9], reticulin and basement membrane [7], the blood group substances in gastric and duodenal mucosa [10], and proteins in muscle [18].

Figure 3 illustrates an interesting new development of the technique. It is a frozen section of a malignant tumour (naevocarcinoma) in human skin and shows specific staining of the normal epidermis by an anti-human-skin serum fully absorbed to remove other anti-human-tissue antibodies and leave only anti-skin activity. Here the serum has failed to stain the underlying cancer cells: by contrast, benign tumours of skin stained as strongly as the adjacent normal epidermis. Such findings support the view that malignant tumours lack specific material present in the normal tissues from which the tumours have arisen. An important concept here is that normal tissues contain antigenic components that are specific for the tissues. P. Vogt [28] has shown by fractionation of rat-liver cells that the organ-specific antigen is confined to the cytoplasmic membranes. Fluorescent antibody studies corroborate this view in the case of rat liver and other tissues, and further suggest that organ-specific antigen is distributed in surface membranes of tissue cells [23]. This observation may have a bearing on some fundamental biological phenomena, for example, the natural mutual adhesiveness of related cells; the normal organization of cells in tissue; and the 'homing effect', in which injected cells tend to settle in the organ of their origin. Such a surface self-marker might be required for recognition and control of normal cells by the body; its absence in cancer cells could be responsible for their uncontrolled malignant growth.

The demonstration of serum antibodies in man and animals is another important use of fluorescent tracing: the serum to be tested is conjugated and applied to suitable preparations of antigens against which antibodies are suspected to be present in the serum. The method permits the diagnosis of recent illnesses due to bacterial or virus infection: the suspected antigen, obtained by culture, is stained specifically by the patient's serum, which may itself be conjugated or, alternatively, located by the 'sandwich' method using conjugated anti-human globulin. This kind of application is rarely required in practice since simpler diagnostic tests are already available, but there are circumstances in which the fluorescence technique may provide information difficult to obtain in any other way. Figure 6 is an example in which human serum antibody to mammalian nuclear proteins is demonstrated by the production of specific nuclear staining in a normal tissue section. Anti-nuclear antibodies, which do not appear to have species specificity, were first demonstrated by the fluorescent antibody technique in 1957 [12]. They have aroused considerable interest because of their association with a group of human diseases of obscure aetiology, now often referred to as autoimmune diseases. It is believed by some, though by no means proven, that antibodies developed by certain people against their own tissues are responsible, at least in part, for clinical manifestations of disease; there is increasing evidence of this in a type of thyroid disease [24]. Apart from its clinical aspects, auto-immunity has important theoretical implications for fundamental immunology [1]. The questions which arise, but are not yet answered with any unanimity, are: what normally prevents animals from producing antibodies to their own cells and tissues, and why does this mechanism sometimes break down?

FUTURE OF FLUORESCENT TRACING

Direct fluorescent protein tracing has been little explored, but it seems safe to predict increasing use of it in biological research. The information it can provide is histologically precise and complementary to the more quantitative information obtainable by radioactive tracing.

Fluorescent antibody tracing is expanding rapidly from microbiological research and applied medical use, into fundamental studies of auto-immunity and of the immunological aspects of species and organ specificity, embryology, genetics, and carcinogenesis. Its increasing use in veterinary and botanical research may be confidently expected. The method is perhaps the most elegant and specific of all the histochemical techniques; its only limitations are those of immunology itself.

ACKNOWLEDGMENTS

The author has pleasure in acknowledging the help received from colleagues in the fluorescence microscopy research team, Dr J. E. Fothergill, Dr K. B. Fraser, and Dr M. G. McEntegart.

REFERENCES

[1] Burnet, F. M. *Brit. Med. J.*, ii, 645, 720, 1959.
[2] Chadwick, C. S., McEntegart, M. G., and Nairn, R. C. *Immunology*, **1**, 315, 1958.
[3] Chadwick, C. S. and Nairn, R. C. *Ibid.*, **3**, 363, 1960.
[4] Coons, A. H., Creech, H. G., Jones, R. N., and Berliner, E. J. *J. Immunol.*, **45**, 159, 1942.
[5] Coons, A. H. *Internat. Rev. Cytol.*, **5**, 1, 1956.
[6] *Idem. Schweiz. Z. Path. Bakt.*, **22**, 693, 1959.
[7] Cruickshank, B. and Hill, A. G. S. *J. Path. Bact.*, **66**, 283, 1953.
[8] Fraser, K. B., Nairn, R. C., McEntegart, M. G., and Chadwick, C. S. *Ibid.*, **78**, 423, 1960.
[9] Gitlin, D., Landing, B. H., and Whipple, A. *J. exp. Med.*, **97**, 163, 1953.
[10] Glynn, L. E., Holborow, E. J., and Johnson, G. D. *Lancet*, ii, 1083, 1957.
[11] Hobson, P. N. and Mann, S. O. *J. gen. Microbiol.*, **16**, 463, 1957.
[12] Holborow, E. J., Weir, D. M., and Johnson, G. D. *Brit. med. J.*, ii, 732, 1957.
[13] Lacy, P. E. *Exp. Cell. Res. Suppl.*, **7**, 296, 1959.
[14] Leznoff, A., Fishman, J., Goodfriend, L., McGarry, E., Beck, J., and Rose, B. *Proc. Soc. exp. Biol. Med.*, **104**, 232, 1960.
[15] Liu, C. *Ibid.*, **92**, 883, 1956.
[16] *Idem. Ergebn. Mikrobiol. Immunforsch.*, **33**, 242, 1959.
[17] Marshall, J. M. *Exp. Cell. Res.*, **6**, 240, 1954.
[18] Marshall, J. M., Holtzer, H., Finck, H., and Pepe, F. *Ibid., Suppl.* 7, 219, 1959.
[19] Moody, M. D., Ellis, E. C., and Updyke, E. L. *J. Bact.*, **75**, 553, 1958.
[20] Moon, H. D. and McIvor, B. C. *J. Immunol.*, **85**, 78, 1960.
[21] Nairn, R. C., Chadwick, C. S., and McEntegart, M. G. *J. Path. Bact.*, **76**, 143, 1958.
[22] Nairn, R. C., Fraser, K. B., and Chadwick, C. S. *Brit. J. exp. Path.*, **40**, 155, 1959.
[23] Nairn, R. C., Richmond, H. G., McEntegart, M. G., and Fothergill, J. E. *Brit. med. J.* ii, 1335, 1960.
[24] Roitt, I. M. and Doniach, D. *Brit. med. Bull.*, **16**, 152, 1960.
[25] Riggs, J. L., Seiwald, R. J., Burckhalter, J. H., Downs, C. M., and Metcalf, T. G. *Amer. J. Path.*, **34**, 1081, 1958.
[26] Schiller, A. A., Schayer, R. W., and Hess, E. L. *J. gen. Physiol.*, **36**, 489, 1953.
[27] Scott, D. G. *Immunology*, **3**, 226, 1960.
[28] Vogt, P. *Z. Naturf.*, **15b**, 221, 1960.
[29] White, R. G. *Nature, Lond.*, **182**, 1383, 1958.

Ultra-high vacuum

G. LEWIN

The phrase ultra-high vacuum has no precise definition, but is generally taken to apply to pressures less than 10^{-11} atmospheres. In recent years the attainment of vacua of this kind has become increasingly necessary, and commercial equipment has been developed. Ultra-high vacua are important in the study of photoelectric, thermionic, and field emission of electrons, and to the physicist interested in the properties of absolutely clean surfaces.

In an earlier article in ENDEAVOUR, E. N. da C. Andrade [1] has recounted the evolution of the vacuum pump, bringing the story up to the early nineteenth century, when pressures of about 20 torr could be obtained. (The torr, named in honour of Torricelli, is the unit of pressure commonly used in high-vacuum physics: 760 torr equals one atmosphere.) Today very much lower pressures are necessary. A pressure of 10^{-6} torr is typical for today's high-vacuum systems used in the production of electronic tubes. Although not precisely defined, the ultra-high vacuum region is generally taken to begin at about 10^{-8} torr, but 10^{-9} torr may be said to be more typical of modern practice.

At pressures such as these, gases are very rarefied indeed, but conditions are still some way from those in a perfect vacuum. The gas-kinetic data indicate how some of the significant properties of air change with diminishing pressure (table 1). In this table the second column shows the number of molecules per cubic centimetre. In their random motion the molecules travel on the average a certain distance (the mean free path) before they collide with another molecule: this distance is shown in the third column. It is clear that, in high vacuum and ultra-high vacuum, molecules collide with the wall rather than with each other.

The fourth column shows the number of molecules that will hit a surface of one square centimetre area each second. If we assume that this surface was initially free from gas and that each molecule that collides with the surface sticks to it, the time required to form a monomolecular layer of gas is that given in column five; a monomolecular layer of gas contains 10^{15} molecules per cm^2. It follows that ultra-high vacuum is required to retain a clean surface for a reasonable length of time. Ultra-high vacuum is therefore of paramount importance wherever surfaces have to be studied and wherever very clean surfaces are desired. Typical applications are the preparation of transistors, and of magnetic and super-conductive films by vapour deposition; and the study of photoelectric, thermionic, and field emission of electrons.

Some characteristics of gas discharges are significantly affected by impurities present at concentrations of as little as one part per million. Studies of such discharges require clean walls and low base-pressures. The need for these experimental

TABLE 1

Gas-kinetic data for air

	1 Pressure (torr)	2 Molecules (cm^3)	3 Mean free path (cm)	4 Molecules incident on cm^2 per second	5 Time to form monolayer (sec)
Atmospheric pressure ..	760	$2{\cdot}6\times10^{19}$	7×10^{-6}	$2{\cdot}3\times10^{23}$	4×10^{-9}
Vacuum	1	$3{\cdot}5\times10^{16}$	5×10^{-3}	3×10^{20}	3×10^{-6}
High vacuum	10^{-6}	$3{\cdot}5\times10^{10}$	5000	3×10^{14}	3
Ultra-high vacuum	10^{-10}	$3{\cdot}5\times10^{6}$	5×10^{7}	3×10^{10}	$2{\cdot}88\times10^{4}$

conditions brought about the systematic investigation of ultra-high vacuum by D. Alpert and his collaborators twelve years ago [2].

Clean surfaces are also required in connection with machines for thermonuclear research, the operation of which is adversely affected by gas impurities, as has recently been described in ENDEAVOUR [3]. Gases used in these devices must be extremely pure, to avoid adsorption of impurities on the walls of the container and their subsequent release under the influence of the discharge.

A new field of application is in the testing of components that have to perform in outer space—such qualities as mechanical friction can be affected by the presence of gas layers on the surface. Also, with space research we need to measure the composition of the atmosphere at great altitude; obviously the equipment itself must contribute as little gas as possible. Besides these immediate applications, it is of interest to the vacuum physicist just to obtain a vacuum as extreme as possible.

MEASUREMENT OF ULTRA-HIGH VACUUM

To investigate ultra-high vacuum it is necessary to have means of measuring it. The commonly used method is the ionization gauge, initially developed by Buckley. It is built like an ordinary triode, but the potentials are very different from those commonly used in radio valves. It consists of a filament surrounded by a grid, which in turn is surrounded by a metallic cylinder. The filament is earthed, the grid has a positive potential of about 150 volts, and the anode is biased 30 volts negative. The electrons which leave the filament are collected by the grid; in transit they ionize the residual gases in the tube. The ions are repelled by the grid because of their positive charge and are collected by the plate (the anode), which is biased negative with respect to the grid. The number of ions produced at a given pressure depends on the type of gas, since the probability that an electron will ionize an atom is a function of the atomic structure.

The highest pressure that can be measured by a conventional ionization gauge is about 10^{-3} torr. If the pressure is reduced below 10^{-8} torr, the ion current remains at a minimum value corresponding to this pressure. Investigators sometimes suspected that they had a pressure considerably lower than 10^{-8} torr—for example, when they observed the emission from points in a field-emission electron microscope—but their ionization gauges nevertheless always indicated this same minimum pressure.

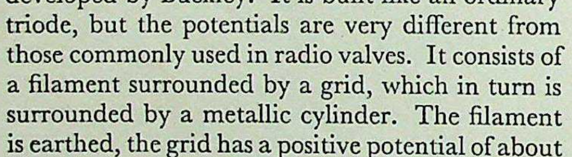


FIGURE 1 – *Bayard-Alpert inverted ionization gauge.*

The explanation of this was first given by Nottingham. The electrons that hit the grid produce soft X-rays. (This phenomenon is employed in X-ray tubes, though there the energy of the electrons, and therefore the energy of the X-rays, is 100–1000 times larger.) These X-rays release photoelectrons from the surface of the ion-collecting plate. The electron current from the ion collector to the grid is electrically indistinguishable from an ion current going to the plate, and is measured in addition to this ion current, and therefore the smallest ion current that can be measured is about equal to this photoelectric current. Its value in conventional gauges corresponds to a pressure of 10^{-8} torr.

In 1947 Bayard and Alpert introduced a gauge, now bearing their names, that reduces this photoelectric current a thousandfold. This is the inverted ionization gauge (figure 1), in which the filament is outside the grid, and the ion collector, consisting of a thin wire only, is inside it. The photoelectric current is proportional to the surface area of the ion collector, and the wire has a considerably smaller area than the cylinder in the earlier gauge. The inverted ionization gauge permits the measurement of pressures down to the 10^{-11} torr range, and is the type commonly used in ultra-high-vacuum work.

One difficulty remains. The heating effect of the hot tungsten filament releases gas from the other parts, and chemical reactions occur between the hot filament and some gases. J. P. Hobson and P. A. Redhead [4] therefore developed a cold-cathode ion gauge which permits measurements of pressures as low as 10^{-13} torr. This gauge is designed like a magnetron. The gas discharge is sustained at low pressures by lengthening the path of the electrons; this increases the probability of an electron hitting a molecule.

All ion gauges act as pumps, since the ions that hit the ion collector or the wall will be buried. This effect is made use of in the so-called ion pumps, which will be discussed later.

Recently, mass spectrometers have been increasingly replacing ion gauges. Mass spectrometers measure the number of atoms of any particular mass present in a given volume of gas; this means that in effect they measure the partial pressure of each gas. Since, as will be shown later, the composition of the residual gas left in an evacuated system has very little resemblance to that of the gas originally present, it is very desirable to determine not only the total pressure but the composition of the gas. Mass spectrometers have no X-ray limitation and they can be built with a high sensitivity, measuring partial pressures as low as 10^{-12} torr. Their disadvantage is that they are considerably more complicated and more expensive than ion gauges.

Mass spectrometers are made in a number of different forms. In one common design, ions are produced in the same manner as in the ion gauge or by a cold-cathode discharge. These ions are accelerated in an electric field and deflected by a magnetic field. For each value of field strength, only ions having a particular ratio of charge to mass can reach the collector, which is located behind a narrow slit.

THE PROCESS OF EVACUATION AND THE DESIGN OF VACUUM SYSTEMS

It is customary to measure the pump action in terms of a pumping speed, S, expressed in litres/sec: S is the volume of gas removed from the container each second at the pressure existing at the time. The reason for this definition is that the pumping speed of most pumps is approximately constant over a large range of pressures. The total mass of gas removed each second is called the throughput of the pump and is expressed in torr litres/sec: it decreases as the pressure drops.

If a pump is connected to a container an equilibrium will be reached eventually, as is shown schematically in figure 2. At the equilibrium pressure P the pump removes the amount of gas SP each second; this is equal to the total influx of gas, which comes from three sources. Q_P is the gas coming from the pump and the pump line, that is, gas desorbed from the wall of the pump line; Q_W is the gas desorbed from the walls of the container; and Q_L is gas leaking into the container from the outside. Q_P and Q_W may decrease with time, but only slowly. In ultra-high-vacuum systems Q_P is often the largest source of gas; in other words, the ultimate vacuum is limited by the performance of the pumping system. The gas initially filling the volume is removed and the gas eventually present differs from that originally present.

Factors affecting Q_P will be discussed in conjunction with pumps. Q_L represents the leaks in the system; though often quite annoying, they can be avoided by proper techniques and the use of a sensitive mass-spectrometer leak-detector. Q_W is the gas given off by the walls, and the development of methods of reducing this gassing rate was the most important step in the attainment of ultra-high vacuum. Gas absorbed in the bulk of a wall, metallic or otherwise, will diffuse to the surface when the pressure is lowered. Metals generally contain a volume of gas which, if reduced to normal temperature and pressure, approximately equals the volume of the metal: this is a considerable quantity. Desorption of this gas by diffusion is very small at room temperature, but it increases rapidly as the temperature is raised. In addition, gas is adsorbed at the surface, by either physical or chemical adsorption. Physical adsorption indicates a rather weak bond between the gas molecules and the surface due to van der Waals forces and having a binding energy of about 1–6 kcal/mol. Physically adsorbed molecules can be desorbed by baking at a few hundred degrees. Chemical adsorption corresponds to a much stronger bond, similar to the ordinary chemical bond. Baking at a few hundred degrees still leaves at least a monolayer of chemically adsorbed molecules.

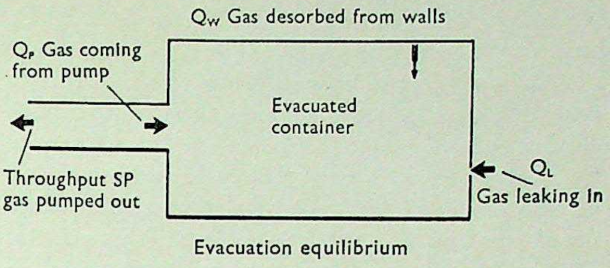

Q_W Gas desorbed from walls
Q_P Gas coming from pump
Evacuated container
Throughput SP gas pumped out
Q_L
Gas leaking in
Evacuation equilibrium

FIGURE 2 – *Schematic representation of factors affecting equilibrium in an evacuated container.*

Although much work has been done, the evolution of gas by desorption is not sufficiently understood to make it possible to predict the behaviour of materials in ultra-high vacuum. Usually only specific examples of outgassing have been published, and from these it is not possible to formulate generally applicable laws. During evacuation, metal and glass surfaces give off water vapour corresponding to many monolayers, besides some carbon monoxide and dioxide. Their initial gassing rate is about 10^{-8} torr litre/cm² sec, but this gradually decreases. H. A. Steinherz [5] showed experimentally, and illustrated by calculations, that, with large stainless steel containers of over 1000 litres capacity, a pressure of 10^{-8} torr can be reached after pumping for one day; if the container is baked for five hours at 260° C, however, a pressure of 2×10^{-9} torr is obtained after a total time of only 12 hours. Baking is thus very effective in reducing the gassing rate. The gassing rate of glass and stainless steel will decrease by a factor of 10^{7} (to 10^{-15} torr litre/cm^{2} sec) by baking at 400° C for 10–20 hours, and this is the usual method for obtaining ultra-high vacuum. Materials that withstand temperatures up to 500° C and have low vapour pressures are therefore used in the construction of the vacuum container. In practice this means glass, metals, and ceramics; organic materials are avoided. Even if the system is not baked, these inorganic materials are preferred because they have lower gas-evolution rates and lower vapour pressures.

Although the value of baking is obvious, to achieve it with complete vacuum systems required a considerable extension of vacuum technology. Alpert designed a small bakable valve made completely of metal; other metal valves have since been developed, and described in the literature. In one design, closure is obtained by pressing a copper nose against the sharp corner of a seat made of stainless steel. The copper flows, and the seat forms an intimate contact with the copper. Figure 3 shows the seat assembly of a valve 20 cm in diameter, designed by J. T. Mark [6]. It employs the 'broaching' principle: the corner of the stainless steel seat shaves off the surface of a copper cylinder and forms its mating surface.

Another important piece of equipment is a demountable joint. It is common in vacuum technique to make joints with 'O' rings that form elastic gaskets between two metal flanges. Most 'O' rings consist of synthetic rubber, but since organic materials have to be avoided in ultra-high-vacuum work, metal gaskets (usually aluminium, copper, or gold) have to be substituted. Several designs have been proposed; figure 4 shows the 'corner gold seal'. In this the gasket is a small gold wire located in a corner between two flanges. The cost of the gold wire is not excessive, since the gold in used gaskets can be recovered.

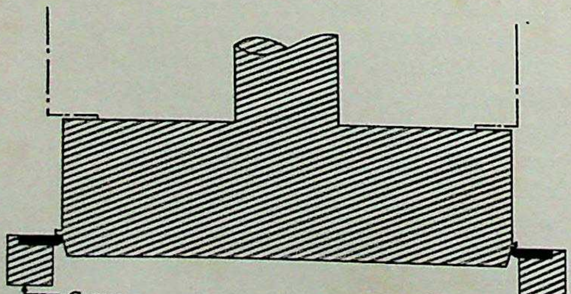


FIGURE 3 – *Seat assembly of valve designed by J. T. Mark* [6].

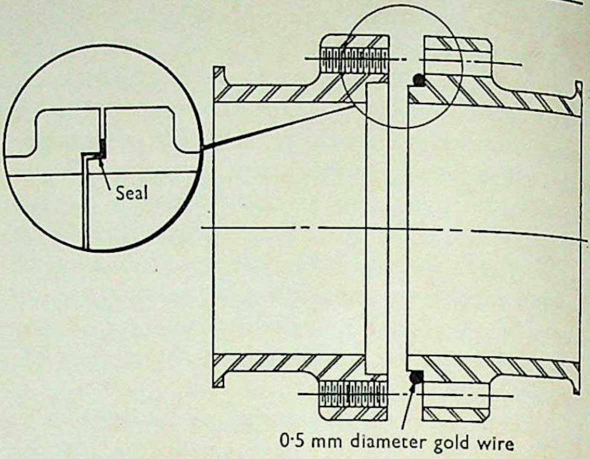


FIGURE 4 – *Ultra-high-vacuum joint with gold-wire gasket.*

Ultra-high-vacuum techniques applied to thermonuclear research have been described by D. J. Grove [7], and numerous literature references can be found in an article by G. L. Mundy [8]. Figure 5 shows an ultra-high vacuum system being tested; the pump box and pump line are seen in the left foreground. The system consists of a 20 cm diameter stainless steel pipe, in the shape of a race track, used in thermonuclear research. It is suspended in such a manner as to permit thermal expansion during baking. The parts are heated by insulating blankets containing built-in heater wires (not shown). The sections are connected by demountable flanges of the type described above.

PUMPING SYSTEMS

So far we have discussed the pumping of ultra-high-vacuum systems without saying what kind of pumps are actually used and what the limitations of these pumps are in terms of gas coming back from them. It must be remembered that the whole pumping system usually consists of a combination of several pumps. Until recently the pump system almost exclusively used consisted of a mechanical fore-pump, a diffusion pump, and a vapour trap, the three being connected in series. All these components had been employed for many years.

The fore-pump contains a rotary piston eccentrically located in a housing and running in an oil bath. Air is sucked in from one side and pushed over to the exhaust side; numerous designs are available. These pumps will pump down from atmospheric pressure to about 10^{-3} or 10^{-4} torr.

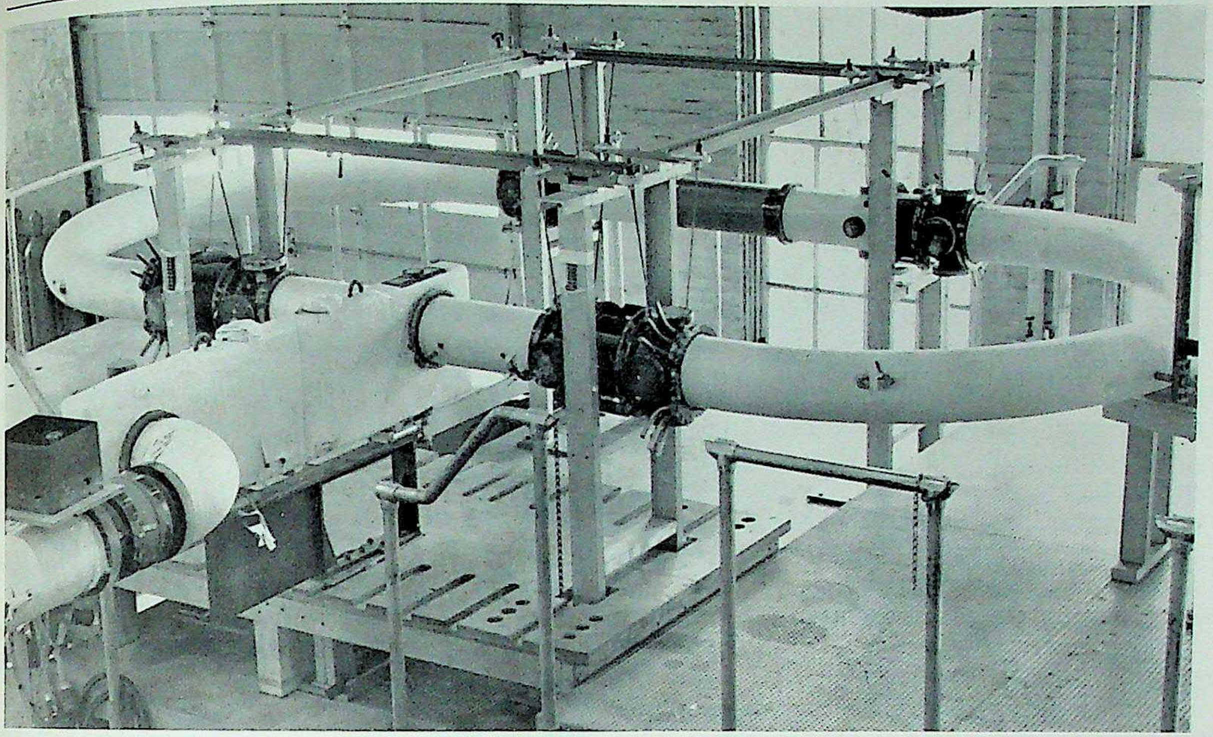


FIGURE 5 – *Ultra-high-vacuum apparatus required for thermonuclear research. The stainless steel tube of which it is constructed is 20 cm in diameter. The pumping apparatus is illustrated in figure* 8.

The lower pressure will be reached only with fresh oil and with two units connected in series.

The diffusion pump contains a liquid, either mercury or a special oil, that is evaporated from a boiler and streams out of a jet to hit the cooled wall of the pump (figure 6). It condenses at the wall and flows back to the boiler. The vapour leaving the jet collides with the gas molecules and gives them an additional momentum in the direction of the fore-vacuum. A pressure difference is thus produced between the space above the jet, that is, the high-vacuum side, and the fore-vacuum. These diffusion pumps can be built in large sizes having pumping speeds of thousands of litres per second.

Since some of the vapour of the pump fluid will be carried over into the high-vacuum side, it is essential to have a trap between this pump and the vacuum system. Without the trap, the limiting pressure would be the vapour pressure of the pump fluid at room temperature, which is 10^{-3} torr for mercury and about 5×10^{-8} torr for diffusion-pump oil. Two types of trap are commonly used; cold traps and adsorption traps. The cold trap shown in figure 7 puts a cold wall in the path of the vapour molecules, and on this the molecules will be condensed. These traps are cooled with liquid nitrogen to $-196°$ C, at which temperature the vapour pressures of mercury, pump oils, and water are extremely small. The trap shown in the picture, which is of a design suggested by Post, has the additional advantage of forming a cold-creep barrier—essential in the case of oil-diffusion pumps—that prevents the oil from creeping along the wall beyond the trap. Such a barrier is not necessary in the case of mercury, because this does not wet the walls and consequently does not creep along them. A cold trap acts as a pump for condensable vapours.

Any clean surface acts as an adsorption trap; the action of activated charcoal, for example, is

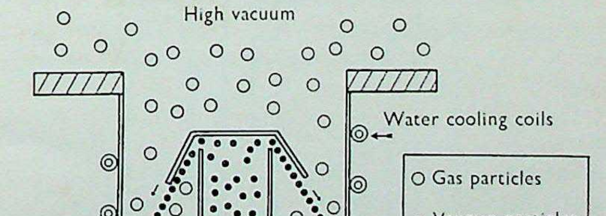


FIGURE 6 – *Schematic representation of mode of action of diffusion pump.*

well known. For ultra-high-vacuum use it must be possible easily to degas the adsorbent. This means that porous bodies, which are the first choice as adsorbers because of their large surface area, must not have internal surfaces that are connected to the outside through small pores only.

No good adsorbent is known for mercury vapour. However, Alpert found that copper foil adsorbs oil vapour, and he built laboratory systems containing a glass oil-diffusion pump with a small copper foil trap (figure 9). After baking the trap and the vacuum system, a pressure of the order of 10^{-10} torr can be maintained for over a month. When the trap becomes saturated, the pressure rises to the vapour pressure of the oil at room temperature, and re-baking is necessary. Biondi showed that traps containing zeolite or activated alumina as adsorbents have a much larger capacity and can also be readily degassed.

Pump oil decomposes to some extent during the heating process. Of the products, carbon monoxide is not removed by a cold trap, but it can be removed by an adsorption trap. Nevertheless, the ultimate pressure obtained in a diffusion pump system with an adsorption trap is not lower than that of a similar system using a cold trap: this pressure is about 10^{-10} torr for mercury and oil pumps alike. Most measurements were made with the Bayard-Alpert ionization gauge, and it appears that this gauge becomes unreliable at this low pressure. With either pump, the remaining gases are mainly carbon monoxide, methane, and helium; the latter diffuses into the system from the outside atmosphere through the glass walls.

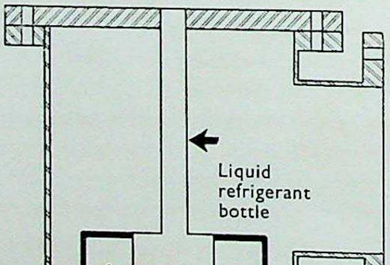


FIGURE 7 – *Trap, cooled with liquid nitrogen, incorporating barrier to prevent creep of oil from the diffusion pump.*

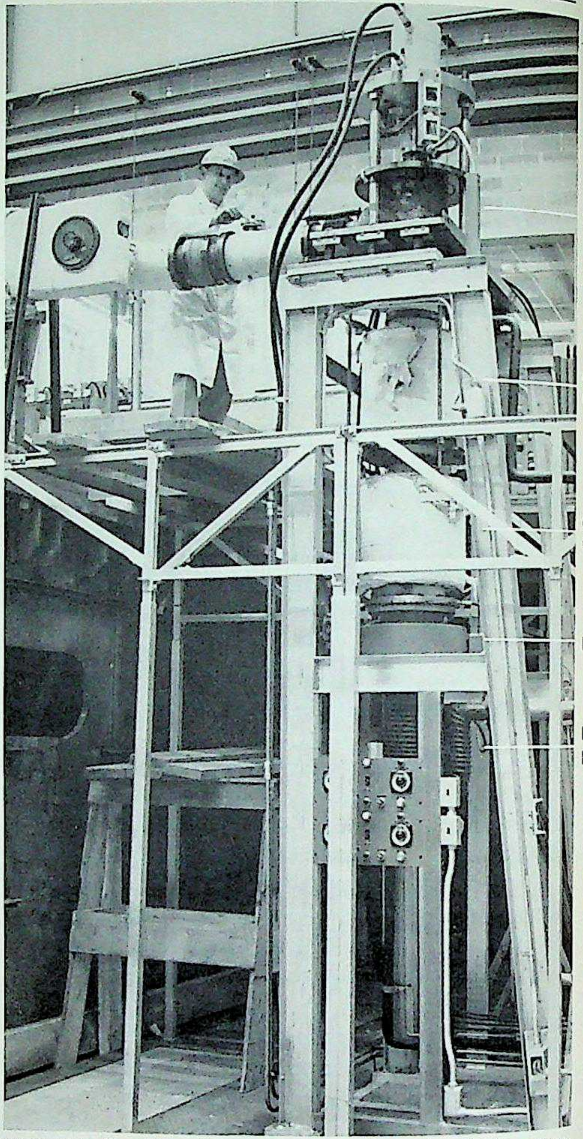


FIGURE 8 – *Pumping apparatus for ultra-high-vacuum system required in thermonuclear research.* (*See also figure* 5.)

Figure 8 shows the pumping apparatus for the vacuum system illustrated in figure 5; the mercury diffusion pump, having a diameter of 25 cm, is at the bottom (the fore-pump is not visible). Above the pump is a refrigerated baffle and two liquid nitrogen traps, which condense the mercury diffusing into the high vacuum, and the bakable valve described previously. The pumping speed is 500 litres/sec. This pumping system maintains a pressure in the 10^{-10} torr range.

The 'pumping' of gas by a film of evaporated metal is used extensively to improve the vacuum

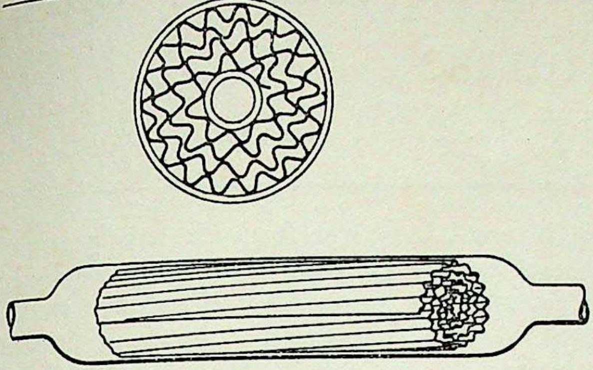


FIGURE 9 – *Copper foil trap for adsorbing oil vapour.*

in electronic tubes: the metal combines chemically with some gases and adsorbs most of them when condensing. The application of this technique to ultra-high vacuum is complicated by the facts that the resulting compounds may have too high a dissociation pressure and that the metals often contain too much gas, which is released during heating. Further, inert gases are not appreciably adsorbed at room temperature. However, it has been demonstrated repeatedly that pumping by evaporated-metal films is possible in ultra-high vacuum if it is preceded by thorough degassing at a high temperature.

Another pumping principle is that of ion pumping, which has already been referred to. In this, gas molecules are ionized, accelerated by an electric field, and buried in an ion collector. Some metal is sputtered off the collector by the ion bombardment; this metal adsorbs gas and so enhances the pumping effect. Such pumps have become increasingly popular for high-vacuum work, but in the ultra-high-vacuum region the pumping speed is greatly reduced.

A special application of ion bombardment is for the cleaning of surfaces by a gas discharge, a procedure used successfully in some thermonuclear machines. The ion bombardment of a surface releases gas molecules adsorbed at the surface and replaces them with the molecules of the bombarding gas. If this gas is an inert gas, such as helium, it is then possible to remove it by prolonged pumping.

Just as pumping by evaporated-metal coatings is an extension of the adsorption trap, so cryogenic pumping—that is, condensing of gases at a cold surface—is an extension of the cold trap. We have found in our laboratory that the pressure in a degassed ion-gauge that is sealed off from an unbaked system drops from the 10^{-7} to the 10^{-10} torr range if a portion of the wall is cooled with liquid nitrogen. The water vapour is rapidly 'pumped' by the cold spot, and the pumping speed of the gauge is large enough (10^{-3}–10^{-2} litres/sec) to cope with the evolution of helium and carbon monoxide. Cryogenic pumping is usually done with either liquid helium (boiling point $-269°$C) or liquid hydrogen (boiling point $-253°$ C). P. A. Redhead [9] obtained pressures in the 10^{-13} torr range by dipping a finger-like extension of the vacuum system into liquid helium. At $-269°$ C hydrogen has a vapour pressure of 10^{-6} torr: apart from helium, it is the only substance that still has an appreciable vapour pressure at that temperature. At $-253°$ C only hydrogen, helium, and neon still have appreciable vapour pressures; the vapour pressure of carbon monoxide at that temperature is less than 10^{-10} torr. D. J. Santeler [10] has analysed this type of pumping and recommends it for large chambers for simulating conditions in space. He suggests the use of panels cooled with liquid hydrogen, these being shielded against the radiation from the chamber by other panels cooled with liquid nitrogen. Russian workers [11] have investigated two liquid hydrogen pumps having pumping speeds of 4000 and 13 000 litres/sec respectively. An advantage of cryogenic pumping is that high pumping speeds can be obtained by dispensing with the connecting pipe between the pump and the vacuum vessel and cooling part of the vessel wall.

REFERENCES

[1] ANDRADE, E. N. DA C. *Endeavour*, **16**, 29, 1957.

[2] Review article by D. ALPERT in 'Handbuch der Physik', Vol. XII, p. 609, 1958.

[3] GOTTLIEB, M. *Endeavour*, **19**, 62, 1960.

[4] HOBSON, J. P. and REDHEAD, P. A. *Can. J. Phys.*, **36**, 271, 1958.

[5] STEINHERZ, H. A. 'Ultra High Vacuum Techniques for Industrial Size Systems'. Transaction New York University Vacuum Metallurgy Symposium. (In press.) 1960.

[6] DREYER, K. and MARK, J. T. 'Ultra High Vacuum System Developments for the Model C-Stellarator'. *Vacuum Symp. Trans.*, 176, 1959.

[7] GROVE, D. J. *Vacuum Symp. Trans.*, 9, 1958.

[8] MUNDAY, G. L. *Nuclear Instr. Methods*, **4**, 367, 1959.

[9] REDHEAD, P. A. *Vacuum Symp. Trans.*, 148, 1958.

[10] SANTELER, D. J. 'Pressure Simulation of Outer Space'. Rep. No. 59GL228, September 1959. General Electric Co., Schenectady, New York.

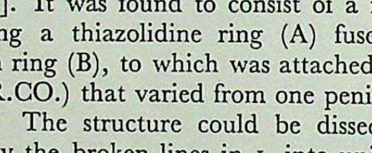
[11] LAZAREV, B. G., BOROVIK, E. S., EYEDOROVA, M. F., and TSIN, N. M. Украïн. Жур. Физ., **2**, 175, 1957.

New penicillins, cephalosporin C, and penicillinase

E. P. ABRAHAM and G. G. F. NEWTON

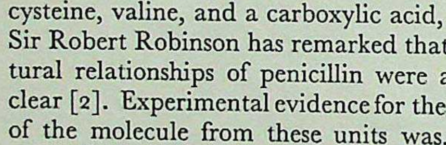
A serious limitation to the value of penicillin in medicine arose from the emergence of strains of *Staphylococcus aureus* that showed resistance to the common penicillins because they produced the enzyme penicillinase. In some hospitals the proportion of resistant strains has been as high as 80 per cent. Later, the staphylococcus was found to acquire resistance to other antibiotics. This article describes advances in the chemistry and biochemistry of penicillin and related compounds that suggest important new medical possibilities.

When the extensive Anglo-American work on the penicillins came to an end in 1946, the problem of producing benzylpenicillin from *Penicillium chrysogenum* on a commercial scale had been largely solved and the pharmaceutical industry was beginning to look for new antibiotics. The general structure of the penicillins (I) had been elucidated [1]. It was found to consist of a nucleus, containing a thiazolidine ring (A) fused to a β-lactam ring (B), to which was attached a side-chain (R.CO.) that varied from one penicillin to another. The structure could be dissected, as shown by the broken lines in I, into units of L-cysteine, valine, and a carboxylic acid, $R.CO_2H$; Sir Robert Robinson has remarked that the structural relationships of penicillin were always sun clear [2]. Experimental evidence for the biogenesis of the molecule from these units was, however, available only in the case of the side-chain. The addition of a variety of mono-substituted acetic acids to the culture fluid of *Penicillium chrysogenum* had been found to increase the yield of antibiotic and to result in the formation of a penicillin whose side-chain was derived from the acid added. For example, the addition of phenylacetic acid ($C_6H_5.CH_2.CO_2H$) greatly increased the yield of benzylpenicillin (I, $R = C_6H_5.CH_2$).

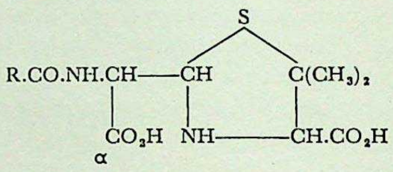


I. *General structure of the penicillins*

$R = C_6H_5.CH_2$ in benzylpenicillin



$R = \overset{+}{H_3N}(\overset{-}{O_2C})CH.CH_2.CH_2.CH_2$ in cephalosporin N



II. *Penicilloic acid formed from penicillin in the presence of penicillinase*

After several years in which little that was new and significant appeared, further advances began to be made. The initial reasons for these advances were largely academic in character, but, at a later stage, activity was undoubtedly stimulated by clinical problems that emerged during the extensive use of penicillin in medicine.

In clinical practice a serious problem arose from a change in the strains of *Staphylococcus aureus* that were causing staphylococcal infections. When penicillin was first introduced into medicine, most strains of staphylococci were highly sensitive to it. As time passed, resistant strains began to be isolated from patients, particularly in hospitals. The proportion of staphylococcal infections caused by resistant strains rose to more than 80 per cent in some hospitals, and the clinical problem was greatly increased by the fact that the staphylococcus showed a tendency to become highly resistant to other antibiotics as well. The staphylococci that were resistant to penicillin were found to produce the enzyme penicillinase, which destroys penicillin. Strains that produced penicillinase became more numerous by selection. They survived and multiplied when the sensitive strains were killed off, and in hospitals they were carried by the medical and nursing staff and passed from patient to patient. Staphylococci that do not produce penicillinase, but are resistant to penicillin for other reasons, can readily be obtained in the laboratory by the repeated subculture of sensitive organisms in the presence of increasing amounts of the drug. But strains of this type have not hitherto presented a clinical problem, perhaps because they are less virulent than their parents. It thus seemed reasonable to hope that a valuable new chemotherapeutic agent would be obtained if the penicillin molecule could be modified in such a way that it retained its high activity but lost its sensitivity to penicillinase.

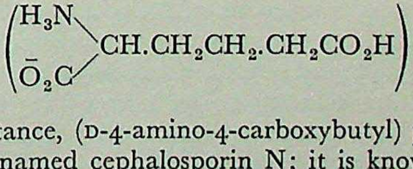
Penicillinase—discovered, in 1940, in the cells of *Escherichia coli* [3]—catalyses the hydrolytic opening of the β-lactam ring of a penicillin with the formation of an inactive penicilloic acid (II). Relatively large amounts of extracellular penicillinase were shown later to be formed by strains of *Bacillus cereus* and it is the enzyme from this source that was first subjected to detailed study. The synthesis of penicillinase by some strains of *B. cereus* proved to be inducible, occurring at a greatly increased rate when the cells came into contact with penicillin [4]. Until recently, staphylococcal penicillinase had received comparatively little attention, but it is now known that the enzyme produced by many strains of *Staphylococcus aureus* is partly extracellular and partly intracellular and that it is inducible. Penicillinases from different sources produce the same qualitative change in the penicillin molecule, but they may differ quantitatively in their enzymic activity.

The hydrolysis of a penicillin by penicillinase may be described in the normal terms of velocity constants for formation and dissociation of an enzyme-substrate complex and for the decomposition of the complex into enzyme and product. With a given concentration of enzyme, the rate of formation of product increases with the concentration of substrate until it eventually approaches a maximum, V_{max}, when virtually all the reactive sites on the enzyme molecules are occupied by substrate. The concentration of substrate required for the rate of reaction to reach half V_{max} is given by the Michaelis constant, K_m, which depends, at least in part, on the affinity of the substrate for the enzyme. With benzylpenicillin as the substrate, V_{max} is high and K_m is low, so that the penicillin is inactivated rapidly even in low concentrations. A useful decrease in sensitivity to penicillinase might thus result from a modification of the penicillin molecule which greatly increased K_m, or from one which greatly reduced V_{max}. In the former case, the new molecule might show too small an affinity for the enzyme to be hydrolysed at a significant rate in low concentrations, although in these concentrations it could nevertheless exert a bactericidal effect. In the case in which V_{max} was reduced, the affinity for the enzyme might still be high. If this were so, the new molecule should not only be resistant to inactivation by penicillinase but should compete for penicillinase with benzylpenicillin and so inhibit the inactivation of the latter.

In this article we shall discuss some of the post-war work in the penicillin field that has now impinged on the problem of the penicillinase-producing staphylococcus. One line of work has made it possible to obtain new penicillins by synthetic or semi-synthetic methods. Another, which developed from a study of antibacterial compounds produced by a species of *Cephalosporium*, has revealed the existence of a substance, cephalosporin C, which is structurally related to the penicillins but does not contain the classical penicillin nucleus.

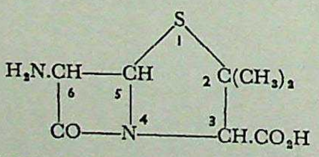
NEW PENICILLINS

By 1950, several hundred different penicillins had been obtained by adding various side-chain precursors, many of them derivatives of phenylacetic acid, to fermentations with *Penicillium chrysogenum*. In general, these substances did not differ greatly from benzylpenicillin in their range of antibacterial activity and they showed no significant resistance to penicillinase.

There was, however, a different antibiotic, produced by a species of *Cephalosporium*, that was later shown to be a penicillin with a new type of side-chain. G. Brotzu had isolated this fungus from the sea near a sewage outfall off the Sardinian coast and reported, in 1948, that a crude extract of its culture fluid exerted a beneficial effect on certain bacterial infections in man. His work was published in a local and little-known journal [5], but it was brought to the attention of Sir Howard Florey by a former British Public Health Officer in Sardinia. Brotzu sent a culture of his organism to Oxford, and further work was done on it both there and at the Antibiotics Research Station at Clevedon. The antibiotic that had first been encountered in Sardinia turned out to be a penicillin with a side-chain derived from D-α-aminoadipic acid  $\left(\begin{matrix} H_3\overset{+}{N} \\ \bar{O}_2C \end{matrix} > CH.CH_2CH_2.CH_2CO_2H\right)$ [6]. This substance, (D-4-amino-4-carboxybutyl) penicillin, was named cephalosporin N; it is known in the United States as synnematin B [7].

The polar groups in the side-chain of cephalosporin N endow the molecule with extremely hydrophilic properties and a range of antibacterial activity that is quite different from that of the common penicillins. The new compound had less than one per cent of the activity of benzylpenicillin against the Oxford strain of *Staphylococcus aureus*, but was considerably more active against certain strains of *Salmonella typhi*. When the positive charge on the side-chain was removed by acylation of the amino group, the activity against the staphylococcus increased and that against the salmonella went down. It thus became clear that radical changes in biological properties could be brought about by certain types of change in the side-chain of the penicillin molecule. However, cephalosporin N was rapidly destroyed by penicillinase from *Staphylococcus aureus* and *Bacillus cereus*, V_{max} being about half that for benzylpenicillin and K_m with the latter enzyme being near that for benzylpenicillin [8]. Cephalosporin N thus showed no promise of being useful for the treatment of infections caused by penicillin-resistant staphylococci.

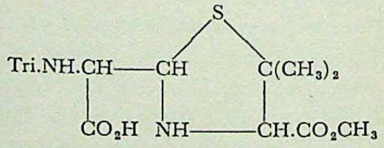
So far, the types of penicillin obtainable had been restricted by the specificity of enzymes involved in the incorporation of the side-chain into the molecule. In general, only monosubstituted acetic acids with non-polar groupings ($R.CH_2.CO_2H$) appeared to be utilized as side-chain precursors by *Penicillium chrysogenum*, while only one penicillin, with a side-chain derived from D-α-aminoadipic acid, appeared to be synthesized by the *Cephalosporium* species. The production of new penicillins by a method to which these apparent restrictions did not apply was made possible by the isolation of the nucleus of the penicillin molecule, 6-aminopenicillanic acid (III).



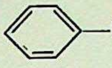
III. 6-*Aminopenicillanic acid* (*the penicillin nucleus*)

The suggestion that the penicillin nucleus was produced in fermentations with *Penicillium chrysogenum* to which no side-chain precursor had been added was first made by K. Kato, in Japan, who found that a substance was present in these fermentations which showed some of the chemical reactions of benzylpenicillin but not its antibacterial activity [9]. In 1950, K. Sakaguchi and S. Murao, also in Japan, had stated that the penicillin nucleus could be obtained by the action of an amidase on benzylpenicillin [10]. The amidase was reported to be formed by a strain of *Penicillium chrysogenum* and to catalyse the hydrolysis of the phenylacetyl side-chain. But difficulty was experienced in repeating these experiments in other laboratories and for some time their validity was doubted.

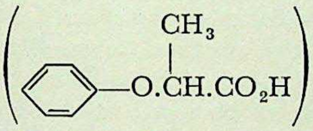
The first independent announcement of the existence of 6-aminopenicillanic acid came from J. C. Sheehan, who reported that the compound had been obtained by total synthesis [11]; this represented the culmination of an extensive series of investigations. Penicilloic acids had been synthesized during the war-time work on penicillin, but attempts to reconstruct the β-lactam ring from them had been frustrated by the ease with which an alternative ring closure, to yield an oxazolone, occurred between the α-carboxyl group and the amide group of the side-chain. The solution of the problem was found in use of side-chains from which oxazolone formation did not occur, together with a new ring-closing agent, dicyclohexylcarbodiimide. 6-Aminopenicillanic acid was finally obtained from a trityl (triphenylmethyl) derivative (IV). After formation of the β-lactam ring, the ester and trityl groups were hydrolysed by alkali and acid respectively under mild conditions [12].



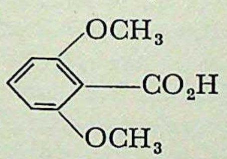
IV

This total synthesis of 6-aminopenicillanic acid made a gratifying addition to the achievements of synthetic organic chemistry, but it involved a number of stages in which the yields were not high and it did not provide a convenient route to the production of large quantities of the compound. 6-Aminopenicillanic acid was first made readily available by the discovery by F. R. Batchelor, F. P. Doyle, J. H. C. Nayler, and G. N. Rolinson, that it could be isolated in crystalline form from fermentations of *Penicillium chrysogenum* to which no side-chain precursor had been added [13]. More recently, amidases that catalyse the hydrolytic removal of the side-chain from penicillins, with the formation of 6-aminopenicillanic acid, have been shown to exist in a variety of micro-organisms [14]. At least two types of enzyme have been detected. One, which occurs in certain fungi and actinomycetes, hydrolyses phenoxymethyl penicillin (I, R = —O.CH$_2$), *n*-amylpenicillin (I, R = C_5H_{11}), and *n*-heptylpenicillin (I, R = C_7H_{15}) more rapidly than it does benzylpenicillin. Another, which occurs in bacteria of the genera *Escherichia* and *Alcaligenes*, hydrolyses benzylpenicillin more rapidly than phenoxymethylpenicillin.

6-Aminopenicillanic acid has a very low antibacterial activity, but it can be readily acylated, for example by treatment with acid chlorides, to yield penicillins, and several thousand new penicillins have already been obtained in this way. As a result of the work of Batchelor and his colleagues, two of these substances have now been introduced into medicine. The first, α-phenoxyethylpenicillin, has a side-chain derived from α-phenoxypropionic acid



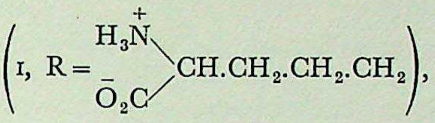
Its maximum rate of hydrolysis by penicillinase is not greatly less than that of benzylpenicillin, but it is useful (like phenoxymethylpenicillin) for oral administration, because it is relatively well absorbed from the gastro-intestinal tract. The second, 2:6-dimethoxyphenylpenicillin, with a side-chain derived from 2:6-dimethoxybenzoic acid



has more interesting properties [15]. It has only about one-fiftieth of the activity of benzylpenicillin against penicillin-sensitive strains of *Staphylococcus aureus*, but its maximum rate of hydrolysis by penicillinase from *Bacillus cereus* is about thirty times less than that of benzylpenicillin and by staphylococcal penicillinase is less still. It induces the formation of penicillinase by both organisms. It is a strong competitive inhibitor of the inactivation of benzylpenicillin by penicillinase from *Bacillus cereus*, but not by penicillinase from *Staphylococcus aureus*, and thus appears to have a much lower affinity for the staphylococcal enzyme. This may diminish its rate of hydrolysis yet further, relative to that of benzylpenicillin, when it comes into contact with penicillinase-producing staphylococci in concentrations no greater than those required to exert an antibiotic effect.

CEPHALOSPORIN C

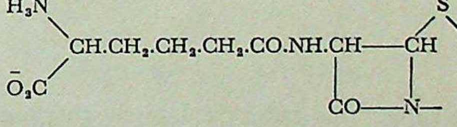
During work on the structure of cephalosporin N



$$\left(\text{I, R} = \begin{matrix} H_3\overset{+}{N} \\ \bar{O}_2C \end{matrix} \rangle CH.CH_2.CH_2.CH_2\right)$$

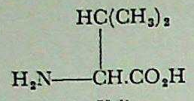
we found that partially purified samples of this penicillin were contaminated by a compound that showed an absorption maximum at 2600 Å in ultra-violet light. The new compound, which was named cephalosporin C, was isolated as a crystalline sodium salt. It proved to have antibacterial activity, but to be only about one-tenth as active as cephalosporin N against most organisms. Cephalosporin C was subsequently shown to be present in culture fluids of the *Cephalosporium* sp. that produced cephalosporin N. At the time of its discovery, its concentration in these culture fluids was too low for it to be detected by measurements of antibacterial activity.

Cephalosporin C was found to contain two more carbon atoms and two more oxygen atoms in its molecule than does cephalosporin N. Like cephalosporin N, it contained a residue of α-aminoadipic acid linked to the rest of the molecule through its δ-carboxyl group and it showed an infra-red spectrum with a strong band at 5·62 μ, characteristic in the penicillin family of the stretching vibration of the C=O group in the fused β-lactam ring. These properties, together with those of a variety of degradation products, indicated that cephalosporin C could be assigned the partial structure V.



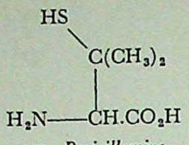
V

This partial structure left seven carbon atoms to be accounted for. Two of these carbon atoms were present as an acetoxyl group ($CH_3CO.O$). The remaining five atoms, like five of the carbons in the thiazolidine ring of the penicillins (I, ring A), formed part of a fragment that yielded valine (VI) on vigorous hydrogenolysis with Raney nickel. Nevertheless, this fragment of cephalosporin C differed from the corresponding fragment of the



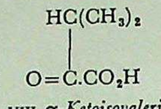
$H_2N\text{—}CH(HC(CH_3)_2).CO_2H$

VI. *Valine*



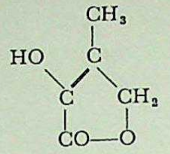
$H_2N\text{—}CH(C(CH_3)_2SH).CO_2H$

VII. *Penicillamine*

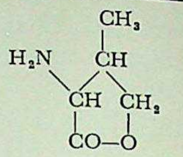


$O{=}C(HC(CH_3)_2).CO_2H$

VIII. α-*Ketoisovaleric acid*



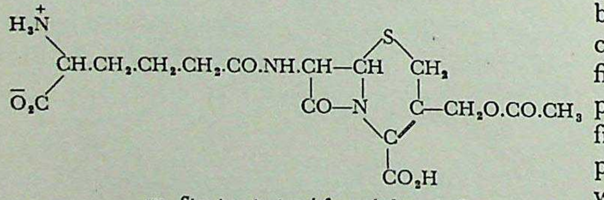
X. α-*Hydroxy*-β-*methylbutenolide*



XI. γ-*Hydroxyvaline lactone*

penicillins, for the penicillins yielded the characteristic amino acid penicillamine (D-β-thiolvaline, VII) on hydrolysis while cephalosporin C did not. During hydrolysis of cephalosporin C, a nitrogen atom was eliminated from the five-carbon fragment in the form of ammonia and when hydrolysis was preceded by mild hydrogenolysis, α-ketoisovaleric acid (VIII) was one of the products. Furthermore, a nuclear magnetic resonance spectrum, determined by P. Higham and R. E. Richards, showed that cephalosporin C, unlike the penicillins, contained no gem-dimethyl group ($C(CH_3)_2$), despite the presence of such a group in two of its degradation products (VI and VIII).

At this stage we proposed for cephalosporin C, as a working hypothesis, the structure IX, in which sulphur was linked to a γ-carbon of the valine-yielding fragment to form a six-membered ring [16]. The proposed structure accounted for the behaviour of the molecule on hydrolysis and hydrogenolysis and for its infra-red and nuclear magnetic resonance spectra. It could not have been predicted that such a structure would show an absorption maximum at as long a wavelength as 2600 Å in ultra-violet light, but no adequate model-compounds were available with which a comparison could be made.



IX. *Structure proposed for cephalosporin C*

Soon after this point had been reached, D. Hodgkin and E. N. Maslen, who had been carrying out an X-ray crystallographic analysis of the sodium salt of cephalosporin C, were able to discern the presence of a six-membered sulphur-containing ring in the molecule [17]. From then on new degradation products were rapidly isolated that gave further support to structure IX. Among them were α-hydroxy-β-methylbutenolide (X) and γ-hydroxyvaline lactone (XI), both of which were obtained after mild hydrogenolysis and subsequent treatment with hot acid.

At the same time, the X-ray crystallographic analysis made rapid progress and soon reached a stage at which the relative positions of the atoms in the molecule were defined, although not with great accuracy. These positions, as shown by the contours representing electron-density levels in figure 1, corresponded with structure IX.

Cephalosporin C thus contains a fused β-lactam-dihydrothiazine ring system in place of the β-lactam-thiazolidine ring system of the penicillins. The positions of the atoms in the β-lactam ring, and of those directly attached to this ring, are very similar in the two structures, although the structures differ considerably in other respects. The dihydrothiazine ring does not appear to have been previously encountered in nature.

Although the chemical properties of cephalosporin C appeared to be sufficiently interesting in themselves to be worth detailed exploration, interest in this antibiotic was enhanced by its biological properties. Its sodium salt was even less toxic to mice than is sodium benzylpenicillin. Its mode of action was similar to that of the penicillins: like the latter, it brought about the lysis of growing staphylococci and appeared to inhibit the synthesis of staphylococcal cell walls [23]. It was a powerful inducer of the formation of penicillinase by *Staphylococcus aureus* and *Bacillus cereus*. In contrast, its maximum rate of hydrolysis was about five thousand times smaller than that of benzylpenicillin in the presence of purified penicillinase from *Bacillus cereus* and also much smaller in the presence of staphylococcal penicillinase. There was no indication that the striking reduction in V_{max} was accompanied by any comparable decrease in affinity for penicillinase from *Bacillus cereus*, since cephalosporin C competitively inhibited the hydrolysis of benzylpenicillin by this enzyme [8]. However, the hydrolysis of benzylpenicillin by extracellular penicillinase from a strain of *Staphylococcus aureus* was not significantly inhibited by cephalosporin C in concentrations similar to those of the penicillin.

Cephalosporin C was as active *in vitro* against penicillinase-producing strains of *Staphylococcus aureus* as against the Oxford strain that did not produce penicillinase. Nevertheless, its activity

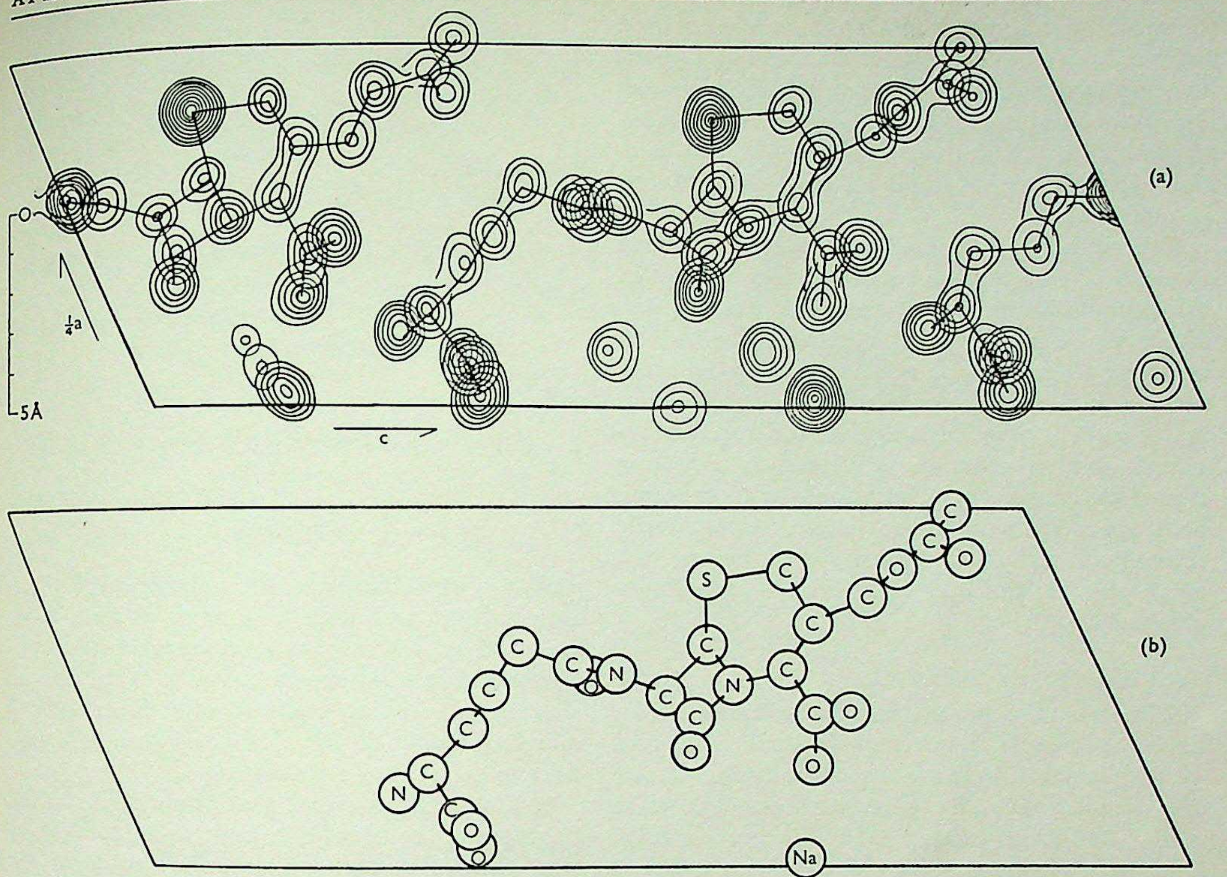
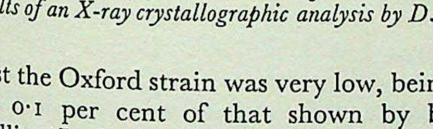


FIGURE 1 – (a) *The structure of the sodium salt of cephalosporin C as it appears from the three-dimensional Fourier synthesis seen down the b axis. The figure represents the asymmetric unit (two molecules in the crystal). The contours represent electron density levels corresponding to the various atoms. Atoms that are not labelled represent the oxygen atoms of water molecules.* (b) *The structure of the sodium salt of cephalosporin C drawn from the atomic positions in* 1a. (*From the results of an X-ray crystallographic analysis by D. Hodgkin and E. N. Maslen.*)

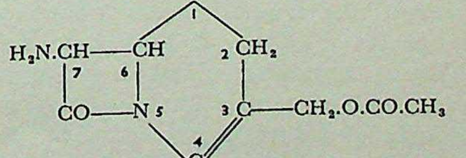
against the Oxford strain was very low, being only about 0·1 per cent of that shown by benzylpenicillin. It thus seemed worth while to explore the possibility of obtaining a derivative of cephalosporin C that retained the resistance of the parent compound to penicillinase but showed a much higher activity against the staphylococcus.

Since benzylpenicillin was more than one hundred times as active as cephalosporin N



$$\left(1,\ R = \begin{matrix} H_3\overset{+}{N} \\ \bar{O}_2C \end{matrix} \!\!> CH.CH_2.CH_2.CH_2\right)$$

against penicillin-sensitive strains of *Staphylococcus aureus*, we thought it likely that a very large increase in activity would result from an exchange of the α-aminoadipyl side-chain in cephalosporin C for a side-chain derived from phenylacetic acid or one of its derivatives. An attempt was made to remove the side-chain from cephalosporin C, by mild acid hydrolysis, without changing the remainder of the molecule. This attempt was successful and although the process was a highly inefficient one, it enabled the nucleus of cephalosporin C, 7-aminocephalosporanic acid (XII), to be isolated in relatively pure form [18].



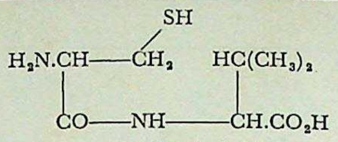
XII. *7-Aminocephalosporanic acid (the cephalosporin C nucleus)*

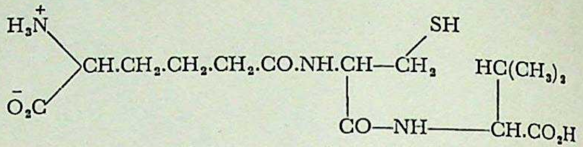
When 7-aminocephalosporanic acid was acylated with phenylacetyl chloride it yielded a phenylacetyl derivative which was several hundred times as active as cephalosporin C, and had about one-fifth of the activity of benzylpenicillin against the Oxford staphylococcus. However, unlike benzylpenicillin, this derivative was highly active *in vitro* against strains of staphylococci that produced penicillinase.

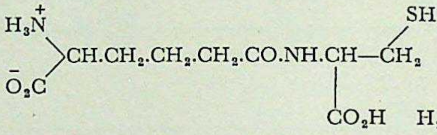
The similarity in the effect produced by these changes of side-chain on antibacterial activity in the cephalosporin C and penicillin families was found to extend further. Penicillins formed by coupling phenoxyacetic acid or α-phenoxypropionic acid to 6-aminopenicillanic acid are at least as active as benzylpenicillin against the staphylococcus *in vitro*, whereas those formed by coupling dimethoxybenzoic acid, propionic acid, or acetic acid are considerably less active. Comparable differences in activity were found among the corresponding derivatives of 7-aminocephalosporanic acid.

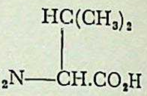
BIOGENETIC RELATIONSHIPS

The structural relationships of benzylpenicillin, cephalosporin N, 6-aminopenicillanic acid, and cephalosporin C indicate that these antibiotics are related biogenetically. However, benzylpenicillin and 6-aminopenicillanic acid are produced by *Penicillium chrysogenum*, whereas the two cephalosporins, with side-chains derived from D-α-aminoadipic acid, are produced by a species of *Cephalosporium*. The differences in enzymic constitution that are responsible for the fact that the two types of fungi produce different, though structurally related, compounds have still be to determined, but there is an indication that α-aminoadipic acid plays a role in the biosynthesis of penicillins produced by *Penicillium chrysogenum* as well as in that of cephalosporin N and cephalosporin C. H. R. V. Arnstein and D. Morris showed that L-cysteinyl-L-valine (XIII) could be used by *Penicillium chrysogenum* for the biosynthesis of benzylpenicillin without prior hydrolysis into cysteine and valine. The dipeptide was not used as effectively, however, as would have been expected had it been an intermediate on the only biosynthetic pathway leading to the antibiotic. A possible explanation of this finding seemed to be provided by the discovery of L-α-aminoadipic acid and δ-(α-aminoadipyl)cysteinylvaline (XIV) in the mycelium of *Penicillium chrysogenum* [19]. The tripeptide (XIV) has a formal resemblance to glutathione (γ-glutamylcysteinylglycine) and might be synthesized in similar stages from its component amino acids. In that case, a major biosynthetic pathway would involve the condensation of α-aminoadipic acid with L-cysteine to form δ-(α-aminoadipyl) cysteine (XV) and the subsequent condensation of the latter with L-valine (XVI). But an alternative pathway would involve the transfer of α-aminoadipic acid from δ-(α-aminoadipyl) peptides to L-cysteinyl-L-valine (XIII).


XIII. *Cysteinylvaline*

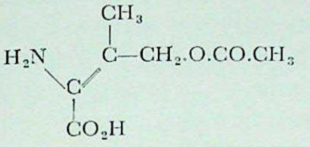

XIV. δ-(*α-Aminoadipyl*)*cysteinylvaline*


XV. δ-(*α-Aminoadipyl*)*cysteine*


XVI. *Valine*

It has been suggested that the cysteinylvaline fragment of the tripeptide (XIV) undergoes oxidative intramolecular condensations to form the penicillin ring system and that enzymes occur in *Penicillium chrysogenum*, but not in *Cephalosporium*, which can bring about, at some stage in the process, the hydrolytic removal of the α-aminoadipic acid residue or its replacement by a residue of phenylacetic acid [19]. If this were so, L-cysteinyl-L-valine (XIII) would not be an obligatory intermediate in the biosynthesis of penicillins and the extent of its incorporation into the penicillin molecule might depend on the extent to which it was converted to δ-(α-aminoadipyl)cysteinylvaline (XIV) by transpeptidation. Moreover, the fact that the α-aminoadipic acid in cephalosporin N has the D-configuration, while α-aminoadipic acid in the mycelium of *Penicillium chrysogenum* is preponderantly L, might have some bearing on the apparent failure of the *Cephalosporium* to produce benzylpenicillin or 6-aminopenicillanic acid. In a number of micro-organisms, enzymes have been found that catalyse the condensation of 6-aminopenicillanic acid (III) with phenylacetic and related acids, as well as the hydrolytic removal of the latter from the corresponding penicillins [14]. It is thus possible that acylation of 6-aminopenicillanic acid represents a final stage in the biosynthesis of benzylpenicillin by *Penicillium chrysogenum*.

The structure of cephalosporin C can be dissected into D-α-aminoadipic acid, L-cysteine, and the O-acetyl derivative of a hypothetical αβ-dehydro-γ-hydroxyvaline (XVII). This third structural unit might be formed by the oxidation of a valine residue in an intermediate that is also involved in the biosynthesis of cephalosporin N.


XVII

These hypotheses, however, remain to be tested. What has been established is the nature of a number of new compounds which form part of the biochemistry of penicillin or may plausibly be assumed to do so. Among them are δ-(α-aminoadipyl)cysteinylvaline (XIV), cephalosporin N, 6-aminopenicillanic acid, and cephalosporin C. But the discovery of such compounds does not carry with it a knowledge of the way in which they are linked in biosynthetic processes.

THE PROBLEM OF THE PENICILLINASE-PRODUCING STAPHYLOCOCCUS

We now know that specific changes in either the side-chain or the thiazolidine ring of the natural penicillins may yield compounds that have antibacterial activity but a greatly decreased sensitivity to penicillinase. 2:6-Dimethoxyphenylpenicillin, in which the penicillin nucleus, 6-aminopenicillanic acid (III), is condensed with 2:6-dimethoxybenzoic acid, has already proved to be clinically useful for the treatment of infections by penicillinase-producing staphylococci [20]. Cephalosporin C (IX), in which the nucleus is 7-aminocephalosporanic acid, has been shown to protect mice from such infections [21], and compounds in which 7-aminocephalosporanic acid is condensed with phenylacetic, α-phenoxypropionic, or similar acids promise to be effective in smaller doses than cephalosporin C itself.

While these are significant advances, it should not be assumed that they have given us ideal chemotherapeutic agents or that they will necessarily provide a permanent solution to the clinical problems now presented by the staphylococcus. 2:6-Dimethoxyphenylpenicillin is intrinsically much less active than benzylpenicillin against *Staphylococcus aureus* and must be injected into a vein or muscle in relatively large amounts. The resistance of the new compounds to penicillinase, though high, is not absolute and it may vary with the source of the enzyme. Penicillinase produced by *Staphylococcus aureus* is not identical with that produced by *Bacillus cereus*, and the latter organism, at least, is itself able to produce enzymes that differ quantitatively in their effect on cephalosporin C. Moreover, the amide linkage of side-chain to nucleus, as well as the β-lactam ring, may be susceptible to enzymic hydrolysis. One possible consequence of bringing the new compounds into general clinical use would be the emergence of staphylococci that produced new penicillinases, or other enzymes, by which these compounds would be readily inactivated. Another would be the appearance, in patients, of staphylococci that had acquired a different type of resistance. The fact that it is the penicillinase-producing organism that has so far impaired the value of penicillin for the treatment of staphylococcal infections provides no guarantee that non-penicillinase producers will not acquire clinically significant resistance to penicillins with new side-chains. The staphylococcus becomes resistant to many other antibiotics *in vivo* without apparently producing enzymes by which these antibiotics are destroyed.

However, the discovery that quite extensive changes may be made in the penicillin molecule without virtual destruction of its antibacterial activity raises the hope that further progress in this field is possible, and that attempts to learn more of the relationship between structure and biological properties among derivatives of 6-aminopenicillanic acid, 7-aminocephalosporanic acid, and related substances will prove rewarding. Antibacterial activity and resistance to penicillinase are not the only properties that are of interest in this connection. For example, a substance with a low activity, but a high resistance to staphylococcal penicillinase, might have additional value if it combined strongly enough with the penicillinase to be a powerful inhibitor of the action of the enzyme on the normal penicillins *in vivo*. A substance with a limited resistance to penicillinase would be more effective if it failed to induce the formation of the enzyme by *Staphylococcus aureus* than if, like benzylpenicillin, it acted as an inducer. The finding of substances with such properties remains a possibility. Both cephalosporin C and 2:6-dimethoxyphenylpenicillin are competitive inhibitors of the inactivation of benzylpenicillin by penicillinase from *Bacillus cereus*. The structural requirements of effective competitive inhibitors appear to be more rigid with staphylococcal penicillinase than with penicillinase from *Bacillus* *cereus*, but the phenylacetyl derivative of 7-aminocephalosporanic acid has some inhibitory action on the hydrolysis of benzylpenicillin by the extracellular staphylococcal enzyme [22]. 2:6-Dimethoxyphenylpenicillin, cephalosporin C, and a number of other derivatives of 7-aminocephalosporanic acid are inducers of penicillinase, but some of them are more efficient inducers than others in relation to their antibacterial activity.

Within the last few years we have learned much about the effects of certain modifications of the penicillin structure on antibacterial activity, maximum rate of hydrolysis by penicillinase, ability to combine with penicillinase, and ability to induce the formation of penicillinase. It appears that a change in one of these properties is not necessarily accompanied by a parallel change in others, and that the behaviour of a compound to penicillinase may vary with the source of the enzyme. We now require more extensive information about the variation of such properties with structure, and a knowledge, in addition, of the structural features of the site of enzymic activity in penicillinases. When these have been obtained, it may be possible to interpret some of the present findings in terms of molecular structure and to predict the structures of further compounds that would have useful biological properties.

REFERENCES

[1] 'The Chemistry of Penicillin' (edited by H. T. Clarke, J. R. Johnson, and Sir Robert Robinson). Princeton University Press, Princeton. 1949.

[2] Robinson, R. 'The Structural Relations of Natural Products'. Oxford University Press, London. 1955.

[3] Abraham, E. P. and Chain, E. *Nature, Lond.*, **146**, 837, 1940.

[4] Pollock, M. R. in 'The Enzymes', Vol. I, p. 619; Vol. IV, p. 269. Academic Press Inc., New York. 1959.

[5] Brotzu, G. *Lav. Ist. Igiene, Cagliari*, 1948.

[6] Abraham, E. P., Newton, G. G. F., Crawford, K., Burton, H. S., and Hale, C. W. *Nature, Lond.*, **171**, 343, 1953.
Newton, G. G. F. and Abraham, E. P. *Biochem. J.*, **58**, 103, 1954.

[7] Abraham, E. P., Newton, G. G. F., Olson, B. H., Schuurmans, D. M., Schenk, J. R., Hargie, M. P., Fisher, M. W., and Fusari, S. A. *Nature, Lond.*, **176**, 551, 1955.

[8] Abraham, E. P. and Newton, G. G. F. *Biochem. J.*, **63**, 628, 1956.

[9] Kato, K. *J. Antibiotics, Japan*, A.6, 130, 184, 1953.

[10] Sakaguchi, K. and Murao, S. *J. agric. chem. Soc. Japan*, **23**, 411, 1950.
Murao, S. and Karshi, H. N. *Ibid.*, **29**, 400, 1955.

[11] Sheehan, J. C. In 'Amino acids and peptides with antimetabolic activity'. Ciba Foundation Symposium (edited by G. E. W. Wolstenholme and C. M. O'Connor). P. 258. Churchill, London. 1958.

[12] Sheehan, J. C. and Henery-Logan, K. R. *J. Amer. chem. Soc.*, **81**, 5838, 1959.

[13] Batchelor, F. R., Doyle, F. P., Nayler, J. H. C., and Rolinson, G. N. *Nature, Lond.*, **183**, 257, 1959.

[14] Rolinson, G. N., Batchelor, F. R., Butterworth, D., Wood, J. C., Cole, M., Eustace, G. C., Hart, M. V., Richards, M., and Chain, E. B. *Ibid.*, **187**, 236, 1960.
Claridge, C. A., Gourevitch, A., and Lein, J. *Ibid.*, **187**, 237, 1960.
Huang, H. T., English, A. R., Seto, T. A., Shull, G. M., and Sobin, B. A. *J. Amer. chem. Soc.*, **82**, 3790, 1960.
Kaufmann, W. and Bauer, K. *Naturwissenschaften*, **20**, 474, 1960.

[15] Rolinson, G. N., Stevens, S., Batchelor, F. R., Wood, J. C., and Chain, E. B. *Lancet*, **ii**, 564, 1960.

[16] Abraham, E. P. and Newton, G. G. F. *Biochem. J.* (In press.) 1961.

[17] Hodgkin, D. and Maslen, E. N. *Ibid.* (In press.) 1961.

[18] Loder, P. B., Newton, G. G. F., and Abraham, E. P. *Ibid.* (In press.) 1961.

[19] Arnstein, H. R. V. and Morris, D. *Ibid.*, **76**, 357, 1960.

[20] Douthwaite, A. H. and Trafford, J. A. P. *Brit. Med. J.*, 687, 1960, and the six following papers.
Brown, D. M. and Acred, P. *Lancet*, **ii**, 568, 1960.

[21] Florey, H. W. *Giorn. Microbiol.*, **2**, 361, 1956.

[22] Crompton, B., Jago, M., Abraham, E. P., and Newton, G. G. F. (Unpublished experiments.)

[23] Abraham, E. P. *Endeavour*, **18**, 212, 1959.

Echolocation by bats

J. D. PYE

Spallanzani established, in 1794, that bats orientate themselves by use of the ears; he presumed that sounds were involved. It has now been shown that bats do in fact navigate, and find their food, by means of an echolocation system using sounds of very high frequency. The process is very rapid, accurate, and resistant to jamming; the author of this article suggests a probable mechanism based on a general theory applicable to nearly all Microchiroptera.

An important part of man's technological progress has involved the development of a variety of devices to aid and supplement his own sensory apparatus. Many of these have acquired great practical importance in helping him to find his way about, especially under difficult conditions. Thus telescopes, night-glasses, and infra-red 'sniperscopes' increase the definition, sensitivity, and spectral response of the eyes, while microphones and hydrophones aid the ears. All these examples are 'passive' in that they rely on signals from external sources, but some other instruments show no such dependence; they radiate energy themselves and create their own signals by interaction with the surroundings. This 'active' group includes radar, asdic (or sonar), searchlights, and even the humble hand-torch. Devices of this type are achieving increased prominence in many fields, and are often used with advantage to augment the passive ones.

It should hardly be surprising, therefore, to find that a similar development has occurred within the sensory apparatus of some animals. Specialization of the sense organs alone is insufficient to allow the full exploitation of certain habitats, and many examples are now known of active sensory systems that produce their own exploratory energy. Some deep-sea fishes may use their luminescent organs to see their way about. Other fish, such as the Gymnotids and Mormyrids, both fresh-water families, have developed electrical methods of orientation with remarkably high sensitivity [1, 2]; pulses of current are produced by the electric organs, and specialized receptors appear to detect impedance changes that distort the field around the body. The blind cave-fish (*Anoptichthys*) uses the lateral-line organs to detect reflections of pressure waves from its own swimming movements. This is really a semi-active mechanism, as the waves are not produced deliberately, but several of the higher vertebrates do produce sounds specially for their acoustic reflections. At least two genera of birds, the Venezuelan oil bird, *Steatornis* [3], and the 'birds-nest soup' swiftlet, *Collocalia* [4, 5], and also fruit bats of the genus *Rousettus* [6–9], can fly safely in complete darkness, guiding themselves by listening to the echoes of clicking noises made with the tongue. Some porpoises and dolphins make noises that are almost certainly used for underwater navigation, and the faculty may well be widespread among the Cetacea [10, 11]. But true echolocation at present appears to be most highly developed in bats of the sub-order Microchiroptera, most of which are insectivorous and nocturnal.

This was the first example of echolocation to be discovered. In 1794, L. Spallanzani established that bats depend for orientation almost completely on hearing and hardly at all on sight. A series of elegant, if cruel, experiments showed that bats with their ears plugged were helpless, but that blinded bats flew normally and were able to catch their prey when released [12]. H. Hartridge suggested in 1920 [13] that an echolocation mechanism using sounds of very high frequency might explain this paradox, and these sounds were detected by G. W. Pierce and D. R. Griffin in 1938 [14].

It has now been established beyond all doubt that this is the means by which bats navigate and find their food, and an account of the experiments performed by many workers over the last twenty years has been published by Griffin [15]. It has been shown that small bats can fly at speed through barriers of vertical wires only 0·4 mm in diameter and spaced only one wing-span apart. They achieve this with a collision rate far below that expected from random movement, and the collisions that do occur are often only slight ones, such as a light brush with the wing tip. When feeding, the bats can catch small insects, such as mosquitoes, at the rate of one every ten seconds for up to 30 minutes [16]. Each capture is accompanied by intensified emission of sound by the bat; only with very noisy prey does the bat appear to listen passively to the insect's wing-beats. The complete manœuvre may be completed in less than half a second, so that the information must be made available to the animal extremely quickly.

Photographs taken by American workers have shown that the bat pursues its prey with great accuracy at close range. If slight errors are made, so that the insect cannot be seized with the mouth, it may be scooped up by the tip of a wing [17]. All this can be accomplished in complete darkness or when the eyes are covered, but when the ears are carefully plugged even a seeing bat in a good light is completely disorientated and blunders into large objects.

The bat is therefore a highly successful self-directed missile that relies on its targets for fuel. But the analogy is not complete, for bats show an extraordinary resistance to jamming. Hunting often occurs with apparent success in highly compromising situations, such as in heavy rain, or among the foliage of trees and hedges. In certain parts of the world, enormous numbers of bats emerge together each night from their roosting places in caves: the background noise and multiplicity of echoes in these conditions would appear to be extremely confusing, yet accidents rarely happen and the interception of prey is continued.

Some species have been forced to fly in an artificial sound-field of 'white' noise whose frequency band covered that used by the bats. Wires of 0·5 mm diameter were avoided almost as well as in the quiet, and well above the expectation for random flight, even when the noise intensity was some 30 000 times greater than the calculated intensity of the echoes [15, 18]. This performance shows an extremely effective rejection of unwanted signals that at first sight appears difficult to explain. However, recent investigations into the 'cocktail party' effect, in which single conversations can be distinguished from a background of noise and chatter, suggest that two ears may be better than one. A model employing two microphones and a non-linear method of correlation has already shown considerable improvement over the theoretical limit for linear devices [19]. Development of this decision theory may help in an understanding of the bat's ability to deal with similar problems.

So far only two families of bats have been investigated very thoroughly, and they appear to have rather different methods of acoustic orientation. Details of the Vespertilionidae are known from the work of R. Galambos, D. R. Griffin, A. D. Grinnell, A. Novick, and others in North America, and of S. Dijkgraaf in Europe, while the European and African Rhinolophidae (horseshoe bats) have been the special study of F. P. Möhres and E. Kulzer. Only these two types can be described here, but other families that have been examined are often similar to these, or are to various extents intermediate in character [20–22].

The orientation sounds used by all Microchiroptera are produced vocally in the larynx. The specialized structure of this organ was described in detail by M. H. A. Robin [23] and by H. Elias [24], although the functional significance of their findings was not apparent at the time. Ossification and fusion of the cartilages has occurred, and the intrinsic muscles are very well developed, especially the cricothyroids, which apply tension to the two pairs of vibrating membranes. The membranes themselves are short and extremely light, permitting the production of very high frequencies. In the Rhinolophidae there are three resonating chambers opening from the trachea, and the glottis is ring-shaped and can be raised to fit into a cartilaginous ring around the internal nares. The Vespertilionidae, with few exceptions, produce their sounds through the mouth and over a fairly wide angle, while in the Rhinolophidae emission occurs through the nostrils and is beamed forwards. This is brought about by the nose-leaf structure, on the snout (figure 3), which acts as a horn and reflector, and by the nostrils, which are a half-wavelength apart, and cause interference laterally. The nose-leaf is capable of some movement, to direct the beam in different directions, and of changes of shape that may alter the pattern of emission to some extent [25].

Both families emit the sounds in discrete pulses that may attain high energy-levels. Recordings made 10 cm from a Vespertilionid show peaks reaching 60–170 dynes/cm^2, roughly equivalent to the level close to a pneumatic road-drill and rather more than an untrained person can produce by shouting. Rhinolophids may produce even higher intensities. The pulses of the Vespertilionidae are variable in duration, according to circumstances, but are generally very short. The typical pulse of a bat flying indoors lasts only 1–4 msec (figure 2, *a* and *b*); some species may produce longer pulses in the open air and shorten them when obstacles or prey are approached. The pulse-repetition rate is variable, and can change from less than 10 per second, when cruising, to

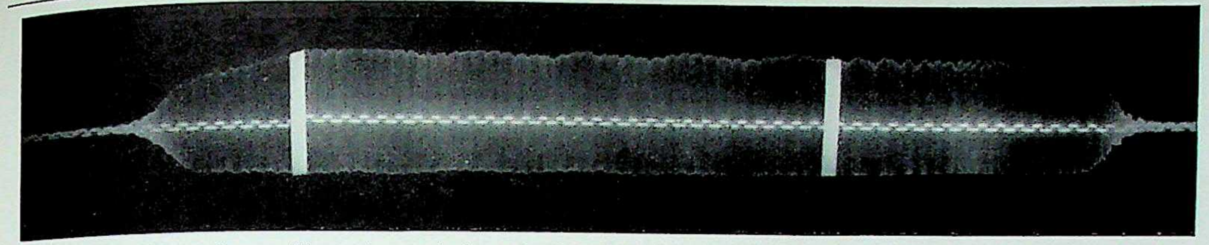
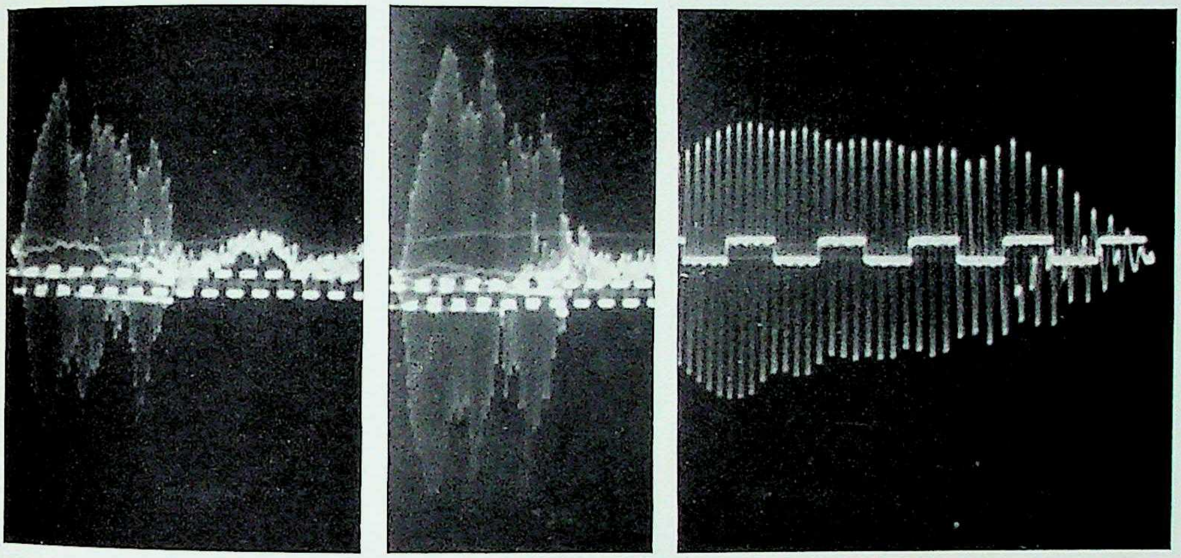


FIGURE 1 – *Part of an oscillograph record of a pulse from* Rhinolophus ferrum-equinum, *the greater horseshoe bat. The time bars are 0·1 msec long, and a total of 54 msec has been removed from the two gaps in the trace. Frequency is very constant but falls in the last 1·5 msec. The trace reads from left to right.*



2a 2b 2c

FIGURE 2 – (a) *and* (b) *Cathode-ray oscillograph traces of pulses produced by* Nyctalus noctula (*Vespertilionidae*) *indoors. The horizontal bars of the time-marking signal are each 0·1 msec in duration.* (c) *The end of a similar pulse expanded to show the continuous fall in frequency. Each horizontal bar represents 0·1 msec. All traces read from left to right.*

over 100 per second when an object is being investigated. The most characteristic feature is that the frequency of the sound falls steadily by nearly an octave throughout each pulse (figure 2*c*). The frequency at the start of each pulse varies, with the species, between 30 and 120 kc/s; shorter, 'close-range' pulses begin at lower frequencies than the longer ones.

In quiet conditions it is sometimes possible to hear a series of faint clicks as a Vespertilionid bat flies about. Each click occurs at the beginning of a pulse, and by listening to these sounds Dijkgraaf was able to determine the rate of pulse production during various manœuvres [26]. At the higher repetition rates, the clicks follow each other so rapidly that they form a distinct buzzing noise whose pitch is easily determined. Each click consists of a few waves of sound at about 10 kc/s, and its intensity is very much lower than that of the subsequent high-frequency sound. Also it is more pronounced in sick, sleepy, or very young bats, and less in those that show the greatest skill in avoiding obstacles. This component therefore appears to play little part in echolocation, but it may be a useful guide to the way the larynx makes the pulses. If the cricothyroid muscles are paralysed by cutting their motor nerves, the bat produces pulses whose frequency remains constant at about 10 kc/s [15]. This may represent the frequency of the chords at their resting tension and subsequent changes in frequency may be in some way produced by muscular action.

The pulses of Rhinolophids are much more constant in form [25]. They last for 40–100 msec and are of very constant frequency, between 85 and 100 kc/s according to the species, although a fall in frequency may be observed during the last few milliseconds (figure 1). Pulse production is co-ordinated with breathing, which is itself synchronized with wing movements in the flying animal. The

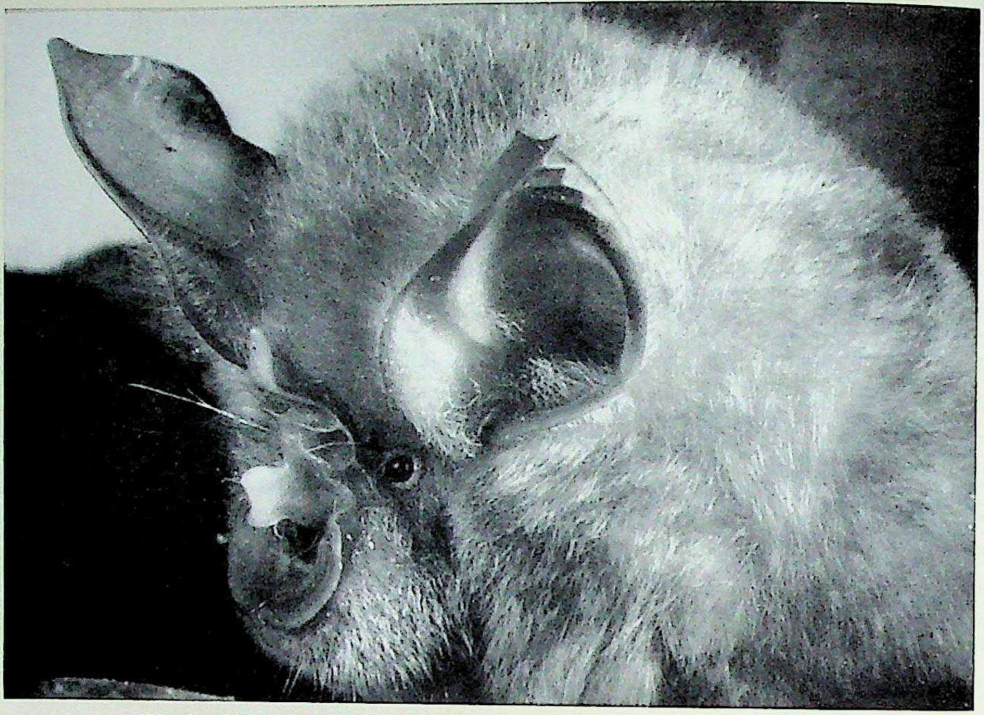


FIGURE 3 – *The head of* Rhinolophus ferrum-equinum (*Rhinolophidae*), *the greater horseshoe bat* (×3). *The nostrils are set close together at the centre of the complex nose-leaf that gives this animal its name. The antitragus is here seen edge-on across the base of the ear.*

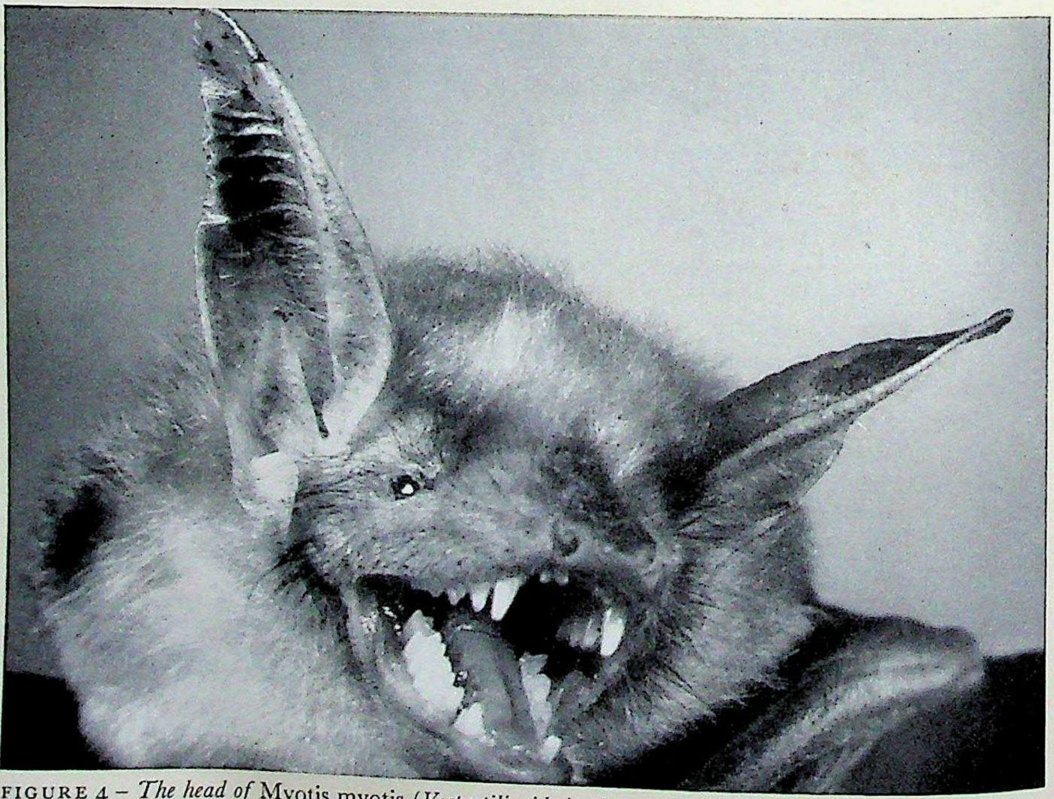


FIGURE 4 – *The head of* Myotis myotis (*Vespertilionidae*), *the mouse-eared bat* (×2·7). *The ears are long but are relatively much smaller than in* Plecotus (*figure* 5).

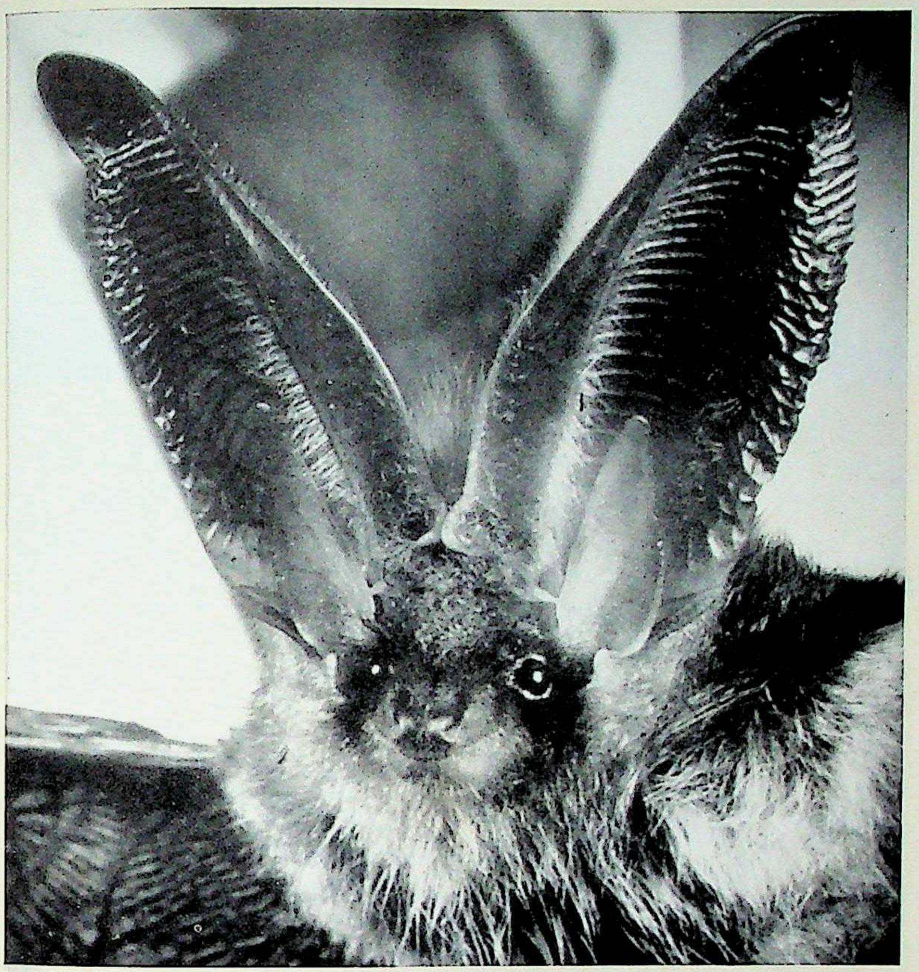


FIGURE 5 – *The head of* Plecotus auritus (*Vespertilionidae*), *the long-eared bat* (×3). *The external ears are enormously developed and are held forwards in flight. The tragus is long and pointed, and the snout has no appendages.*

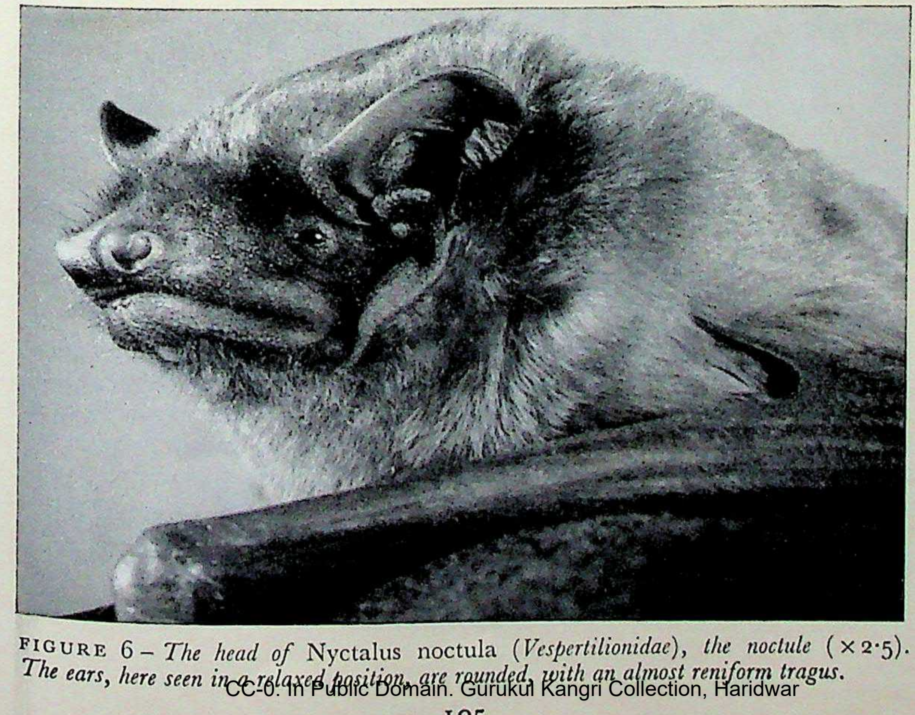
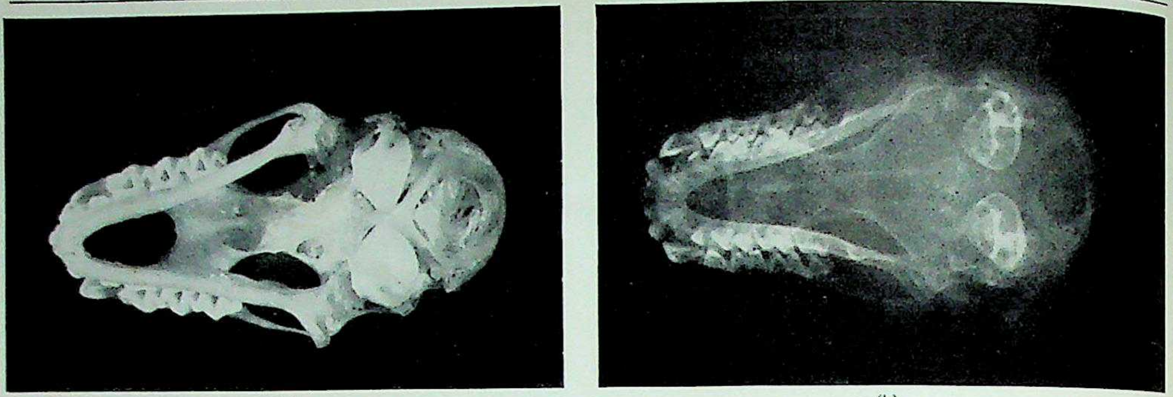


FIGURE 6 – *The head of* Nyctalus noctula (*Vespertilionidae*), *the noctule* (×2·5). *The ears, here seen in a relaxed position, are rounded, with an almost reniform tragus.*



(a) (b)

FIGURE 7 – (a) *Ventral view of the skull of* Rhinolophus ferrum-equinum (×2·6). *The large proportions of the cochleae are apparent, but the bullae which enclose the middle ear are restricted to rings of thin bone on the outer side of each cochlea.* (b) *An X-ray photograph of the same skull to show the full extent of the cochleae.*

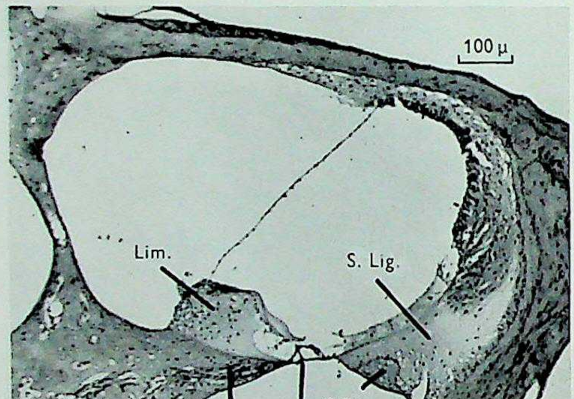


FIGURE 8 – *A photomicrograph of a section of one turn of the cochlea of* Rhinolophus ferrum-equinum. *b.m. = basilar membrane; Lim. = limbus; P.S.L. = primary spiral lamina; S. Lig. = spiral ligament; S.S.L. = secondary spiral lamina.*

pulse-repetition rate therefore seldom exceeds five or six per second in flight, and is not so variable with behaviour as in the Vespertilionids. The pulses of both these families show a very low harmonic content, although other bats may produce marked second and third harmonic components [20].

The ears of bats are highly modified for the reception of high-frequency signals. The external ear is always large—its length may exceed that of the body (figure 5)—but its structure varies considerably. The Vespertilionidae have a fairly simple immovable lobe, with a well-developed tragus looking like a second, smaller ear in front of it. Several suggestions have been made concerning the function of this 'earlet', but so far none has been substantiated. The shape of the tragus usually resembles that of the pinna (figures 4–6), which suggests that it plays some part in determining the acoustic properties of the ear.

The Rhinolophidae have no tragus, but another lobe, the antitragus, is developed as a fold in front of the pinna and forms a funnel around the ear canal (figure 3). The two lobes do not meet at their bases on the inner side: there is thus a second opening to the meatus which faces forward and downward along the snout. The ears of these bats are freely and independently mobile and, when the animal is awake, are kept in constant motion by complex musculature on the head. They appear thus to be actively searching for the sources of sounds or echoes, whereas the Vespertilionidae can turn only the whole head. Möhres [25] has shown that, in addition to these searching movements, the ears are also vibrated forwards and backwards at rates up to 50 times per second, the tip of the ear often moving through an arc of 8–10 mm. This unusual action only occurs while the bat is emitting exploratory pulses, and it may be responsible for a low-pitched buzzing sound that is produced at this time. It seems to be an important part of the echolocation mechanism, since, if the ears are rendered immobile by denervating the appropriate muscles, the bat is disorientated in flight. Later the animal learns to compensate by performing rapid movements of the whole head, and the ability to orientate is regained to some extent [28].

The muscles of the middle ear are very well developed. These are the tensor tympani and stapedius, responsible for tensioning the chain of ossicles that conducts sound from the ear-drum to the inner ear. The bulla is composed of very thin bone and is restricted in extent to the external side of the cochlea (figure 7*a*). The cochlea is very large and occupies much of the posterior part of the skull, especially in the Rhinolophidae, where the two receptor organs almost meet in the mid-line (figure 7, *a* and *b*). The basilar membrane is very narrow and appears to be under tension, since its supporting structures are very well developed (figure 8). On the inside are a thick spiral lamina and a very high limbus; to the outside the spiral ligament is extremely large and is further strengthened by a second spiral lamina of bone. Many features of this receptor are not yet understood, but they represent an extreme form of the pattern usually associated with hearing in the high-frequency ranges [29]. This conclusion is supported by a certain amount of experimental evidence. Cochlear microphonic potentials have been obtained from Vespertilionids as far as 98 kc/s, the limits of the apparatus used [30]. In other experiments, bats were trained to expect food when sounds of various pitch were heard, and responses were obtained up to 200 kc/s [31]. By comparison, hearing in man seldom extends above 17–20 kc/s.

Although the total size of the brain may be extremely small, the auditory regions of the bat's brain are extraordinarily well developed [15]. The visual centres, by contrast, are greatly reduced. Grinnell and Griffin [32] have recorded changes in electrical potential at several levels in response to sound stimuli but especially at the posterior colliculus. They found that the greatest responses are evoked by short pulses of sound and that pairs of pulses can be resolved when separated by as little as 1–5 msec. Inhibitory interaction between signals from the two ears was observed when the sound was projected from certain directions. There was little evidence of noise rejection at this level, but sharply tuned responses were obtained to brief tones as high as 150 kc/s.

These high frequencies are an important feature of echolocation, as they increase the resolution that can be achieved. In general, waves are reflected specularly only by objects whose dimensions are very much greater than one wavelength. As size decreases, the received energy becomes reflected more widely, until objects that are very small compared with the wavelength scatter incident energy equally in all directions. Wavelength is inversely proportional to frequency, and the sounds used by bats, in the range 25–120 kc/s, have wavelengths of 15–3 mm. They are thus able to produce sharp echoes from quite small objects, but the successful detection of targets smaller than this shows that to some extent useful echoes may be obtained from scattered signals.

The range at which detection may occur has been estimated by several methods. Using the increase of pulse-repetition rate of Vespertilionids as a criterion of detection, Griffin [15] has estimated that barriers of wire may be detected indoors at about 2 m, although no avoiding action was taken until the bats were much closer. The same bats in the open will dive 5–6 m to investigate small objects thrown into the air. However Dijkgraaf and Möhres estimated, by a series of training experiments, that accuracy of discrimination extends to a range of only about 50 cm [31, 25]. Rhinolophid bats, on the other hand, appear to perceive their surroundings accurately to a range of about 6 m, although resolution probably decreases with distance.

A further difference between the two families is seen when the ears are temporarily incapacitated. All bats are completely disorientated if the external ears are carefully plugged on both sides, but a Rhinolophid bat is able to fly perfectly well with one ear free, and to avoid obstacles with ease. In strange contrast, the Vespertilionid bat deafened on one side only is loth to fly, and when forced to do so it blunders about as clumsily as when totally deafened.

To summarize, the Vespertilionidae produce many short, frequency-modulated, pulses through a wide angle; their ears are not mobile, and both must be functioning. The Rhinolophidae make fewer, longer, pulses of very constant frequency, and concentrate the sound into a beam with which they scan their surroundings. The ears move independently in a complex manner; orientation may be achieved monaurally; and the bats can probably probe accurately to greater ranges.

How do these mechanisms operate? What information do the bats obtain about nearby objects? Are there two or more distinct methods, or are they variations of a single one? These fundamental questions cannot yet be answered, but several theories have been proposed in attempts to explain the mechanism behind the observed phenomena. The first, that of Hartridge, resembled the principles of early radar in many ways [33]. He supposed that the bat can estimate the range of its targets by measuring the time delay before an emitted sound returns as an echo. Normal binaural location of the direction from which the echo comes will then give precise information about the relative position of the object concerned. The very short pulses of the Vespertilionidae were considered to be suitable for such a mechanism, as they contain sharp peaks of energy that could be used to mark short intervals of time; the form of Rhinolophid pulses was not then known.

Under close examination, several points of difficulty arise regarding this early theory [34]. The time delays involved are very short, and their measurement must be extremely accurate to explain the skill observed. A central mechanism capable of this accuracy must be envisaged. It is known that the time of arrival of a single sound at the two ears can be compared with great accuracy, but the timing of short intervals by the same ear demands a different type of central analyser. Yet it would seem that range is one of the most important pieces of information, and echo delay its most direct means of measurement.

Again, it is difficult to understand how the bat can hear very faint echoes so soon after it has made such a loud noise itself. It is unlikely that much attenuation can be effected within or around the head of such a small animal, and the mammalian ear generally experiences a rise in threshold following a very loud sound. Hartridge suggested that contraction of the middle-ear muscles by the intra-aural reflex may damp the ear during pulse production and then restore sensitivity for echo reception. But the very high speeds of reflex action and muscle relaxation needed make such an idea untenable. Direct motor control of these muscles, synchronized with those of the larynx, might achieve the required result, but there is no evidence or precedent for such an arrangement.

Furthermore, at the very close ranges involved in the interception of insects, the echo front must return before production of the pulse is completed. This overlap would be expected to mask the fainter echo-signal. Griffin has suggested that frequency modulation of the pulses may overcome this difficulty, since the echo delay will ensure that pulse and echo are never heard at the same frequency at the same time, but the difference must be small at very close ranges. The problem could also be reduced if the bat employs the principle of the recently developed chirp-radar [35]. This system uses frequency-modulated pulses similar to those of the Vespertilionidae, and the receiver incorporates a delay network whose effect varies with frequency. If the pulse front is delayed more than subsequent parts, the emergent pulse is compressed in duration, and overlapping echoes may be separated. Is it physiologically possible for the bat's ear to contain such a frequency-sensitive delay network?

The ability to cope with multiple echoes demands that the timing device, whatever its mode of operation, should be able to handle a series of intervals simultaneously and to discriminate between them. The extent to which some bats can deal with such situations would place very great, though not necessarily insuperable, demands on such a system.

Möhres has rejected the idea of delay-time measurement for the Rhinolophids because of the long duration and low repetition-rates of the sounds made by these bats. There are no sharp energy-peaks in these pulses, so that only the relatively infrequent fronts and trailing edges of pulses could be used for timing, and chirp-radar principles cannot apply. Instead, Möhres suggests that the bats observe the loudness of echoes while scanning with their beam of ultrasound and with their presumably highly directional ears. The bearing of echo targets is then easily obtained from proprioceptive information and the range could be obtained by a process of triangulation, although the base-lines involved are very small. This method could operate monaurally and therefore satisfies the observations on this point.

The problem of masking is here even greater than for Vespertilionids. The sound does not change frequency, and the duration of each pulse will ensure that overlap of pulse and echo occurs up to a range of several metres. It is hard to believe that the bat could hear any but the loudest of echoes, and minute observation of their intensities would be very difficult.

A general theory applicable to nearly all Microchiroptera has recently been advanced by the author [34] in an attempt to overcome some of these problems. It is rather striking that the maximum ranges at which each type is thought to locate objects accurately are those at which their respective pulses and echoes just overlap. Possibly the overlap, which would seem to mask at least part of each echo, may be a necessary part of the detection mechanism. It is suggested that under these conditions the bat may not attend to the echo itself; instead, he listens to beats between the echo and the original call. If beats occur at a

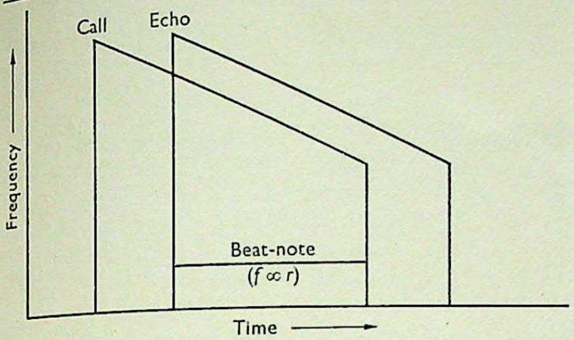


FIGURE 9 – *A diagram of the possible construction of a beat-note from a Vespertilionid pulse and its echo. The beat frequency* (f) *is directly proportional to the range* (r) *of the target.*

frequency greater than the lower limit of hearing, they could be heard as a separate note, a beat-note.

For the frequency-modulated pulses of the Vespertilionids, the frequency of the beat-note will be characteristic of the echo delay and therefore of the range of the target (figure 9). Frequency analysis by the cochlea could thus provide an accurate method of measuring the very short time-intervals involved. Since the frequency of the pulse seldom changes by more than an octave (a factor of two), the beat frequencies will nearly always be below those of the pulse and echo, so that problems of masking will be much reduced. Multiple targets may be distinguished, as each echo produces a separate beat-note and the cochlea is known to be able to discriminate between simultaneous notes of different frequencies. A comparison of beat-notes produced in each ear could also give directional information.

The same idea has been put forward independently by L. Kay [36]. He further suggests that an extension of range could be achieved if the bat repeats each pulse to itself after a controlled delay and so produces beats with later echoes. Without a provision of this kind, one must assume that the bats that detect wires at distances greater than half the pulse-length in air must be using another mechanism. But detection of an object does not necessarily indicate that the bat possesses exact knowledge of its position. Possibly little information is gained from the discrete echoes heard at first, and accuracy is obtained only when the range is reduced and overlap occurs. The agility of bats, which can turn in 30 cm or less, makes it very difficult to judge how much information they have when 2 m from an obstacle. Nevertheless, the bats' initial directional perception of insects appears to be very good [15].

The beat theory applied to Rhinolophid pulses provides a sensitive method for the detection of Doppler shifts (figure 10). The bat flying towards an obstacle could measure the approach velocity, as this is proportional to the beat frequency. But, probably more important, it provides a method of preventing masking. At the relatively high speed of 10 m/s, a pulse of 100·00 kc/s (as used by *Rhinolophus hipposideros*) will return as an echo of 106·25 kc/s and so produce a beat-note of 6·25 kc/s. This is probably the maximum value that will be encountered, and the original pulse will hardly interfere with its detection. At the same time the mechanism is extremely sensitive to the very low relative velocities. Location of the target may then be achieved by observing the loudness, not of the echo as Möhres suggests, but of beat-notes from it and the call.

For this method to work, the ears of the Rhinolophid must be highly directional, and here the very rapid vibration of the pinnae may help. If each pinna acts as a reflector for collecting echoes, its own movement will introduce a range of Doppler shifts which will, at any instant, be proportional to the velocity of each part of the collecting surface. Thus the ear vibrating at 50 c/s will produce a range of beat-notes (figure 11). The 100 c/s component of this mixture will predominate, as it will synchronize with the ear movements. The effect will be greatest for an echo returning in the direction of ear movement and will decrease to either side. Even the stationary bat can detect stationary objects by this means, as the velocity component is supplied artificially. Of course, if the original pulse from the nostrils were heard after reflection from the moving pinna, it would experience Doppler shifts similar to those of the echo and no particular

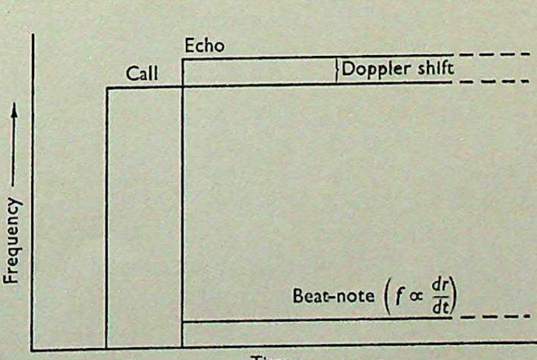


FIGURE 10 – *A diagram showing the production of a beat-note from a Rhinolophid pulse. The beat frequency is proportional to the rate of change of range* (dr/dt), *that is, to the relative velocity of bat and target.*

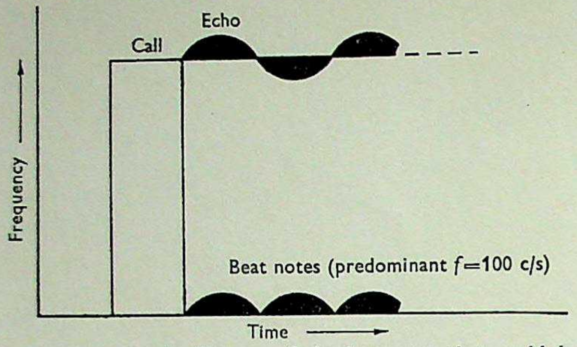


FIGURE 11 – *A diagram of the beat-notes that could be produced by ear vibration in a Rhinolophid. The component whose frequency is twice that of the ear movements (or higher even multiples) will remain coherent and so will predominate. The beat-frequency scale and Doppler-shift scale are here exaggerated for clarity.*

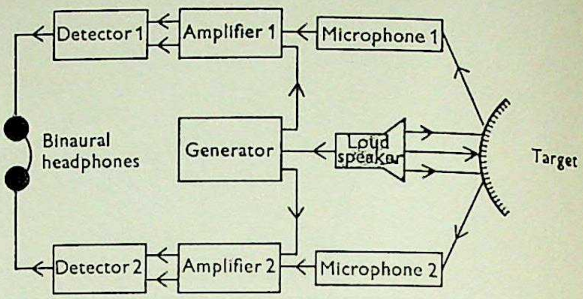


FIGURE 12 – *A block diagram of the apparatus used to test the ability of the human ear to detect objects by beat-note echolocation. The generator produces different types of bat-like sounds, and beat-notes produced with the echoes are heard through headphones. The 'vibrating ear' reflectors for each microphone have been omitted from the diagram for simplicity.*

beat-note component would predominate. It is therefore suggested that the second opening to the ear canal, at the non-moving base of the pinna, may act as a pathway for the 'direct' signal. As may be seen in figure 3, the ventral aperture faces along the snout towards a deep groove in the periphery of the nose-leaf 'transmitter'.

One of the methods of detecting beat-notes is to add the two original signals together and then to subject them to non-linear distortion. The first action occurs within the meatus, and there is evidence that the second may occur within the ear. Studies of the cochlear microphonic potentials of cats and guinea-pigs have shown [37] that strongly non-linear characteristics appear when the ear is overloaded at intensities of 1–10 dynes/cm². In view of the very much higher intensities produced by bats, it would be surprising if stray sounds from the mouth or nostrils did not reach these levels at the ear. Instead of this signal being an embarrassment by masking echoes, it may be an essential part of the mechanism. The site of distortion within the ear is the source of some controversy, but there is some evidence that the type of distortion best suited to beat-note detection is caused by contraction of the middle-ear muscles [38]. This may explain their well-developed condition in the bat.

Normally beat-notes are not very loud to the human ear, as the degree of distortion present is low, but under optimum conditions, that may well occur in the bat's ear, the beat-note amplitude can approach that of the smaller signal, in this case the echo. In order to see how well the human ear can discriminate the position of objects, a model has been built as shown in figure 12. Distortion is provided by a detector in conjunction with artificially generated bat-like sounds, and the operator is easily able to detect objects in any of the three modes proposed above. Another model using the same principle, but differing in detail, has been constructed by Kay, and similar results have been obtained with it.

The beat-note theory supports rather than invalidates the conclusion already reached by other workers. It attempts only to suggest a method whereby some of the expected results may be obtained and many of the drawbacks avoided. By proposing a common physiological basis for the different systems observed it allows a wide range of intermediate conditions and facilitates speculation about the evolution of these systems. It throws no light, however, on the signal-to-noise problem.

Whether the bats do in fact use one of the many methods suggested so far, or whether they have other more subtle means at their disposal, can only be decided by further investigation. This promises to be very exciting, and a better understanding of the bat's speed, accuracy, and resistance to jamming may well be of interest in fields far removed from the study of small mammals.

REFERENCES

[1] LISSMAN, H. W. *J. exp. Biol.*, **35**, 156, 1958.
[2] LISSMAN, H. W. and MACHIN, K. E. *J. exp. Biol.*, **35**, 451, 1958.
[3] GRIFFIN, D. R. *Proc. nat. Acad. Sci., Wash.*, **39**, 884, 1953.
[4] NOVICK, A. *Biol. Bull., Wood's Hole*, **117**, 497, 1959.
[5] MEDWAY, LORD, *Nature, Lond.*, **184**, 1352, 1959.
[6] NOVICK, A. *J. exp. Zool.*, **137**, 443, 1958.

[7] Griffin, D. R., Novick, A., and Kornfield, M. *Biol. Bull., Wood's Hole*, **115**, 107, 1958.

[8] Möhres, F. P. and Kulzer, E. *Z. vergl. Physiol.*, **38**, 1, 1956.

[9] Kulzer, E. *Ibid.*, **43**, 231, 1960.

[10] Schevill, W. E. and Lawrence, B. *Breviora, Mus. comp. Zool. Harvard*, **53**, 1, 1956.

[11] Kellogg, W. N. *J. acoust. Soc. Amer.*, **31**, 1, 1959.

[12] Dijkgraaf, S. *Isis*, **51**, 9, 1960.

[13] Hartridge, H. *J. Physiol.*, **54**, 54, 1920.

[14] Pierce, G. W. and Griffin, D. R. *J. Mammal.*, **19**, 454, 1938.

[15] Griffin, D. R. 'Listening in the Dark.' Yale University Press. 1958.

[16] Griffin, D. R., Webster, F. A., and Michael, C. R. *Anim. Behav.*, **8**, 141, 1960.

[17] Webster, F. A. (Unpublished observations.)

[18] Griffin, D. R. and Grinnell, A. D. *Science*, **128**, 145, 1958.

[19] David, E. E. and Kaiser, J. F. Reported in *Wireless World*, **66**, 467, 1960.

[20] Griffin, D. R. and Novick, A. *J. exp. Zool.*, **130**, 251, 1955.

[21] Novick, A. *Ibid.*, **138**, 81, 1958.

[22] Möhres, F. P. and Kulzer, E. *Verh. dtsch. zool. Ges., Erlangen*, **59**, 1955.

[23] Robin, M. H. A. *Ann. Sci. nat.*, **12**, 1, 1881.

[24] Elias, H. *Morph. Jb.*, **37**, 70, 1907.

[25] Möhres, F. P. *Z. vergl. Physiol.*, **34**, 547, 1953.

[26] Dijkgraaf, S. *Experientia*, **2**, 438, 1946.

[27] Griffin, D. R. *Ibid.*, **7**, 448, 1951.

[28] Möhres, F. P. *Symp. Zool. Soc. Lond.*, **3**, 57, 1960.

[29] Reysenbach de Haan, F. W. *Acta oto-laryng. Stockh.*, suppl. 134, 1958.

[30] Galambos, R. *J. acoust. Soc. Amer.*, **14**, 41, 1942.

[31] Dijkgraaf, S. *Acta physiol. pharm. néerl.*, **6**, 675, 1957.

[32] Grinnell, A. D. and Griffin, D. R. *Anat. Rec.*, **134**, 574, 1959.

[33] Hartridge, H. *Nature, Lond.*, **156**, 490, 1945.

[34] Pye, J. D. *J. Laryng.*, **74**, 718, 1960.

[35] Klauder, J. R., Price, A. C., Darlington, S., and Albersheim, W. J. *Bell Syst. tech. J.*, **39**, 745, 1960.

[36] Kay, L. *Anim. Behav.* (In press.)

[37] Wever, E. G. and Lawrence, M. 'Physiological Acoustics'. Princeton University Press. 1954.

[38] Newman, E. B., Stevens, S. S., and Davis, H. *J. acoust. Soc. Amer.*, **9**, 107, 1937.

Book reviews

RELATIVITY

Relativity, the General Theory, *by J. L. Synge. Pp. xv + 505. North-Holland Publishing Co., Amsterdam.* 1960. £5 10s. *net.*

Books on general relativity differ in their approaches to the geometrization of gravitation that is made possible by the principle of equivalence, that is, by the curious fact that all bodies fall equally fast in a gravitational field. Some books emphasize the gravitational aspect and make the geometry subsidiary: most take a middle way in which geometry and physics are balanced. I do not think there has ever been such a completely geometrical exposition as the one given by Professor Synge. It is true that the important physical tests are referred to (partly in footnotes), and there is a short reference to the recent check of the theory by the Mössbauer effect; but the major portion of the book and its whole spirit are geometrical.

The exposition employs a new method—it is based on Ruse's function, which is essentially the square of the distance between two events in space-time. This is a sign that we are beginning to be able to describe relativity in terms of concepts that bear some resemblance to those involved in classical physics, where it is described as action at a distance.

The book contains accounts of much recent work has so far been available only in the original papers. There is also a bibliography of over 60 pages, listing work published from the birth of the theory until the present time.

C. W. KILMISTER

ELEMENTARY-PARTICLE PHYSICS

Progress in Elementary Particle and Cosmic Ray Physics, *Vol. V, edited by J. G. Wilson and S. A. Wouthuysen. Pp. xii + 461. North-Holland Publishing Co., Amsterdam.* 1960. 85*s. net.*

In 1952, when the first volume of the series 'Progress in Cosmic Ray Physics' appeared, the chief sources of most elementary particles were the high-energy fluxes from cosmic radiation. In the preface to the third volume (1956) it was noted that the operation of the great accelerating machines had brought a new and dominant approach to much of the experimental work on strange particles, and it had to be decided whether elementary-particle physics—as distinct from pure cosmic-ray physics—should be excluded from subsequent volumes. The editors concluded that this was not desirable, and the two fields of interest are now combined in the new title of the fifth volume of the series, which contains reviews on weak interactions by A. Lundby; phenomenology of nucleon-nucleon interaction by J. Gammel and R. Thaler; theory of anti-nucleons by J. McConnel; observations on cosmic ray jet-interactions in nuclear emulsions by D. H. Perkins; and absorption and decay of negative muons by R. M. Tennant.

Admirable as all the articles are, I personally found the first two contributions particularly rewarding reading. These two articles occupy nearly half the volume, and both are timely, as certain aspects of the subjects treated are currently coming to a definitive close. The determination of a unique set of proton-proton-scattering phase shifts up to energies around 300 MeV appears nearly complete, representing the end of a quest that began some twenty-five years back, and it is proper that a review should appear from two authors who have been among the most active in this quest. Similarly, β-decay of the neutron and the muon seems well understood, and although a number of review articles on weak interactions have been published in recent years, these have been written mostly by theoretical physicists. Lundby's, coming from a distinguished experimental physicist, supplies the right emphasis in regard to the 'experimental nuances' of the subject and is altogether welcome.

This is perhaps the place to mention that I still think there is need for a third type of review article on weak interactions—an article tracing the history of the complex interrelation of experiment and theory, telling the tangled story of how false theories led to, dare one say, false experimental findings in the last few years, and *vice versa*. Grodzins (*Proc. Nat. Acad. Sci.*, **45**, 399, 1959) and Blackett (Rutherford Memorial Lecture, 1958) have made a start in this direction but there is still need for a longer review, which would inevitably be charged with many moral lessons. Ideally it should be written by someone who is at once a great theorist as well as a great experimental physicist. There just are not enough writings of this sort in scientific literature. A. SALAM

MATHEMATICAL PHYSICS

The Theory of Brillouin Zones and Electronic Structure in Crystals, *by H. Jones. Pp.* 268. *North-Holland Publishing Co., Amsterdam.* 1960. 60*s. net.*

It is most important, in theoretical physics, to distinguish between exact and approximate mathematical results. The problem of calculating the energy of an electron moving in the periodic field of a crystal lattice has been central to the theory of solids for the past thirty years, and has still not been completely solved. But there are a great many perfectly precise statements one can make without ever solving the equations of motion, from a study of the symmetry properties of the crystal structure. Some elegant theorems can be established from the geometry of the lattice and the formal algebra of group theory. Before one embarks upon a numerical computation—and, again, when one has reached the other shore—one must check against these theorems and make sure they are fully satisfied. Professor Jones has made this task much easier by setting out the whole theory in simple practical terms, in simple practical prose, complete with 66 tables, 94 figures, and indeed almost all the detailed results that one normally needs to know. To the professional mathematical physicist, this book is a valuable instrument, as much so as any spectroscope or bubble chamber. J. M. ZIMAN

PHYSICAL CHEMISTRY

Theory of Unimolecular Reactions, *by N. B. Slater. Pp. ix*+230. *Methuen & Co. Ltd, London.* 1959. 36*s. net.*

This is the first book to be devoted entirely to unimolecular reactions and it gives a full account of the very valuable contributions that the author has made to this subject during the last twenty years. In his treatment, a polyatomic molecule is regarded as made up of a number of classical harmonic oscillators, the motions of which are described in terms of a series of internal co-ordinates, one or more of which must attain a certain critical value for reaction to take place. With the aid of this model—provided that the necessary detailed spectroscopic study and vibrational analysis have been made of the molecules concerned—the author is able to predict, approximately, the frequency factors of unimolecular reactions and, somewhat more successfully, the variation of the velocity constant with reactant concentration. In addition to describing in some detail what is probably the most satisfactory and realistic treatment of the subject yet devised, the author discusses fairly and sensibly the relationship to previous work of his own approach, which he regards not as supplanting but as supplementing and reconciling the older viewpoints.

This is a highly specialized work, and by no means all chemists will want to read, let alone possess, this book. However, all those who are interested in the mechanisms of chemical reactions, and who possess the necessary patience and mathematical background to follow the author's detailed arguments, will find here a valuable and thought-provoking account of an approach to kinetic problems that is of great fundamental importance. C. F. CULLIS

FREE RADICALS

Stabilization of Free Radicals at Low Temperatures. Summary of the NBS Program, *edited by A. M. Bass and H. P. Broida. Pp. iii*+110. *National Bureau of Standards Monograph* 12. 1960. $1.50.

A research group to carry out work for three years on trapped free radicals was set up in the National Bureau of Standards on September 1956. This group, which was under the direction of H.P. Broida, and was generously supported, provides a unique example of a research team. It was housed in the Bureau, but many of the workers came on temporary leave from industry, and several were from outside the United States. It was engaged in work in a field which it was initially thought might provide a new, powerful, rocket fuel, and yet there was no secrecy. Those who came into the group were free, to a very considerable extent, to pursue their own ideas. This book provides, in eight articles, a good ninety-page account of the work done in the three years. There is also a summary of the symposium that marked the conclusion of the programme. It is true that the work showed that trapping atoms and radicals was most unlikely to provide a rocket fuel. However, it did add greatly to our scientific knowledge in many fascinating and provoking ways. Those who are interested in what such a group could and did achieve will find this book valuable. The articles are most informative, and there is a complete list at the end of the book of the papers published by this group. J. W. LINNETT

GRAPHITE

Graphite and its Crystal Compounds, *by A. R. Ubbelohde and F. A. Lewis. Pp. xii*+217. *Clarendon Press, Oxford; Oxford University Press, London.* 1960. 35*s. net.*

There is great technical and scientific interest in graphite. The technical interest arises from its extremely high melting point, good conductivity, and attractive nuclear properties; the scientific from graphite's place as the limit of the various aromatic structures and from its being a layered macromolecule.

A great deal of work has been carried out on graphite but much has been wrapped in secrecy, because of manufacturers' reluctance to reveal production techniques and because of the importance of graphite in atomic-energy projects. This book is most welcome, as it surveys the literature of the past century, with over eleven hundred references to work on the crystallography, chemistry, and physics of graphite.

The chemistry of graphite is that of a crystal containing structural defects. The term 'crystal compound' emphasizes that this is a different sort of chemistry from the classical, since, in many of its reactions, the graphite lamellar skeleton is preserved throughout. It is clear that intensive study of the various crystallographic models of graphite and of the structural defects in these, dealt with in the opening chapters of the book, will lead to a better understanding of the electric and magnetic properties, and chemical reactions, so well surveyed in the later chapters.

A. KELLY

INORGANIC CHEMISTRY

Nouveau traité de chimie minérale, *Vol. XV*, Uranium et transuraniens, *Part I*, Uranium, *edited by P. Pascal. Pp. l+734. Masson et Cie, Paris.* 1960. 115 *NFcs.*

Volume XV, part I, of this treatise deals exclusively with uranium; the forthcoming part II, will be devoted to the transuranium elements. The emphasis in the volume under review is preponderantly on those subjects that are pertinent to nuclear technology: uranium mineralogy, separation of the element from its ores, preparation of the metal and alloys, analysis of uranium-containing materials, and so on. The general chemistry of uranium in solution is covered only briefly. The treatment of major subjects is comprehensive, well organized, and reasonably up to date; most bibliographies extend well into 1958. Figures are numerous, clear, and well chosen. The sections dealing with the alloy systems of uranium and with the physical and mechanical properties of the metal are especially comprehensive, and a chapter devoted to the aqueous corrosion of uranium and its alloys is of considerable technological interest.

Although exception may be taken to some of the more fundamental points of view relating to uranium chemistry —for example, classification of uranium as a member of group VIb of the periodic system—there is no question but that, on the whole, the sub-editors of this volume, R. Caillat and J. Elston, and Professor P. Pascal, the editor-in-chief of this monumental chemical survey, are to be congratulated for their efforts in producing an extremely valuable reference work in the field of modern inorganic chemistry.

G. T. SEABORG

HETEROCYCLIC CHEMISTRY

An Introduction to the Chemistry of Heterocyclic Compounds, *by R. M. Acheson. Pp. xiv+342. Interscience Publishers Inc., New York; Interscience Publishers Ltd, London.* 1960. 35*s. net.*

This admirable book, written by a university teacher and active research worker, gives a clear, concise, and comprehensive account of the main heterocyclic types, including not only relatively well known compounds such as pyrrole, furan, indole, pyridine, and pyrimidine, but also less familar compounds such as azirine, aziridine, azetidine, and quinolizine. Each chapter presents the facts in a logical manner, describing the important physical and chemical properties of representative compounds and, when appropriate, discussing their theoretical significance; considerable attention is devoted to derivatives of biological and technical interest. There are numerous references to review articles and relevant papers, and there is a good subject and substance index. The format and formulae are excellent, and typographical errors are few and for the most part unimportant. The book is most readable and can be strongly recommended to both students and research workers.

N. CAMPBELL

Pyridine and its Derivatives, *edited by E. Klingsberg, Part I, by R. A. Barnes, F. Brody, and P. R. Ruby. Pp. x+613. Interscience Publishers Inc., New York; Interscience Publishers Ltd, London.* 1960. £18 7*s. net.*

This is a part of the fourteenth volume in the valuable series of monographs appearing under the general title 'The Chemistry of Heterocyclic Compounds'. It contains two chapters, written by different authors and very different in content. The second chapter has two parts; one gives a very detailed account of the occurrence of the pyridine ring in naturally occurring compounds and the methods by which these derivatives have been separated, the other an equally detailed account of the methods by which the pyridine-ring system has been synthesized in the laboratory. Much of the material is tabulated. This chapter is for experts, and it will be used as a reference section; the material is clearly set out and easily followed.

The first chapter occupies only 97 pages and is very different in character. It succeeds in conveying a lucid yet comprehensive description of the chemistry of pyridine derivatives. This is no catalogue of chemical reactions. Physical and chemical properties are described and, as far as possible, explained in the light of modern theory. This chapter is extremely readable. It is of interest both to the specialist and to the Honours student who requires a short, intelligible account of the chemistry of this typical hetero-aromatic system. Altogether a very useful volume.

P. F. HOLT

ORGANIC CHEMISTRY

The Alkaloids: Chemistry and Physiology, *Vol. VI* (*Supplement to Vols. I and II*) *by R. H. F. Manske. Pp. xii+442.* $14. *Vol. VII* (*Supplement to Vols. II, III, IV, and V*) *by R. H. F. Manske. Pp. xiii+559.* $17. *Academic Press Inc., New York and London.* 1960.

Alkaloid chemistry is one of the traditional fields of endeavour of the organic chemist, and it continues to provide challenging problems of wide interest. The series entitled 'The Alkaloids', edited by the distinguished authority R. H. F. Manske, in part collaboration with H. L. Homes, has provided during the last decade an excellent summary of the field. As the articles are contributed by many authors, the standard attained is not uniform, but on the whole Manske has chosen his authors well. In any case, the series constitutes a unique reference-source for natural-product chemists.

Like other branches of organic chemistry, the subject is no sooner defined than it proceeds to expand at an exponential rate. Supplementary volumes are therefore always urgently needed, and the two books under review serve this purpose for the five volumes already published. It is perhaps unfortunate that definite dates for the literature coverage are not specified, but literature up to about 1957 or 1958 seems to be covered. The present reviewer would suggest that in future volumes all authors should state exactly to what year and month the literature has been thoroughly searched. Even if such dates were embarrassing to the publishers they would be very helpful to the public. Indeed, the publishers might thereby be shamed into more efficient and speedy printing.

The reviewer can recommend these two books as worthy successors to the volumes that have gone before. They will be indispensable not only to libraries, but also to all interested in the chemistry of natural products.

D. H. R. BARTON

GENETICS

Introduction to Quantitative Genetics, *by D. S. Falconer. Pp. ix+365. Oliver and Boyd Ltd, Edinburgh.* 1960. 35*s. net.*

The theory concerning continuous inheritance cannot as yet be used with as much precision in experimental design and in forecasting results as can that concerning discontinuous inheritance, on which it is based. But to the applied geneticist and to many other biologists it is none the less of more interest, for characters of medical and anthropological importance, and plant and animal characters of economic importance, are in general quantitative.

A book, therefore, that not only synthesizes the very diverse and mathematically advanced work of contemporary writers but also avoids their controversies, presents a uniform notation, and uses only elementary mathematics, should be very welcome to research workers in these fields. The latter feature will appeal particularly to the student; no calculus, matrix algebra, or path coefficients appear. An elementary knowledge of diploid genetics, and of statistics (chiefly analysis of variance, correlation, and regression), is all the reader requires.

The first five chapters describe the properties of populations with single-locus segregation; the next five study the concepts arising from extension to multiple loci; these lead to three chapters on selection and three on inbreeding and cross-breeding; and finally there are four chapters on scale, threshold and correlated characters, and natural selection; possibly linkage could have been more extensively treated. The book is meticulous in exposition and well illustrated with experimental material.

M. E. WALLACE

PLANT TAXONOMY

Taxonomy of Flowering Plants, *by C. L. Porter. Pp. xii+452. W. H. Freeman & Co., San Francisco and London.* 1960. *44s. net.*

Plant taxonomy, the first branch of botany, is relatively unfashionable these days, and a good textbook that will interest university students in the subject is sorely needed. There are adequate elementary books and good advanced manuals; none, however, really suitable for university courses. It is particularly disappointing, then, that this book fails to fill the gap. The first few chapters, on principles, nomenclature, methods, literature, and terminology, are quite satisfactory, although experimental methods might have been discussed in more detail—most of the recent advances in taxonomy have been made by their use. The systematic section is the weak point of this book. It is impossible in a text of this type to deal fully with a very large number of families; this is the function of advanced manuals. The result is that no family is dealt with really adequately. A much better plan would have been to deal with selected families in more detail, to give an idea of the range of variation within, as well as between, families. A good deal more material could have been included had the layout of the book not been so extraordinarily wasteful. Fully 15 per cent of the systematic section consists of blank space; some pages have less than one-third of their space occupied. The quality of many of the full and half-page photographs leaves much to be desired; they could have been profitably replaced with text.

This book is one of a series that contains many excellent textbooks. It is difficult to understand why the editors should have passed this one in its present form. It cannot be recommended, although, in the absence of better books, it is usable.

S. R. J. WOODELL

PLANT PHYSIOLOGY

Plant Physiology, *edited by F. C. Steward. Vol. Ia,* Cellular organisation and respiration. *Pp. xxvii+331. Academic Press Inc., New York.* 1960. $13.

This, the first part (though the second to appear) of a treatise extending to six volumes, contains three chapters. The subjects of two of them are indicated by the sub-title, and the third is devoted to a survey of proteins and enzymes. The chapters are all written by authors with well-established reputations in the fields covered. The treatment is as consistent throughout as can reasonably be expected, and as up to date as the hazards of publication allow in a rapidly moving subject. Moreover the aim is not to give a picture of the fleeting moment but to say what plant physiology is about, and to put it into a reasonable context of morphology and other associated sciences. The aim is also to give a reasoned analysis of the various branches of the subject rather than an exhaustive compilation, or an overall picture. It is, however, claimed that the treatment is sufficiently detailed to benefit the research of specialists and comprehensive enough for the use of students. It would perhaps be rather unkind to consider in detail how far the published chapters achieve two such disparate ideals: they naturally pursue a middle course. In a book intended at least in part for those in training it is unfortunate that the phraseology is rather often slipshod and ambiguous, and that misprints liable to mislead the uninitiated have not been wholly eliminated. Nevertheless, enough has now appeared to make it clear that the treatise as a whole is likely to perform a valuable service in botanical education.

W. O. JAMES

PHOTOPERIODISM

Photoperiodism in Plants and Animals. Proceedings of the Conference on Photoperiodism, October–November 1957, *edited by R. B. Withrow. Pp. 922. American Association for the Advancement of Science, Washington; Bailey Bros. and Swinfen Ltd, London.* 1959. £6 *13s. net.*

Night and day impose diurnal rhythmicity of photosynthetic and other activities; alternation of the seasons requires devices such as dormancy and migration to ensure survival. These seasonal changes are, in many organisms, effected by photoperiodism (sensitivity to the pattern of exposure to light and darkness). An inborn rhythmicity, with a period of approximately 24 hours, is also a common biological property, and the rhythm phase is usually determined by light.

A symposium to survey the facts and theories of photoperiodism was held in October 1957. Its programme seems to have been devised in the hope that diurnal rhythmicity might account for photoperiodic responses to unnatural light-cycles. No individual is likely to possess all the facts available on, and relevant to, photoperiodism, so that bringing all the data together is a most desirable objective.

From the title and bulk of this book one might expect coverage of the whole field, but this is hardly achieved. It deals first with pigments, then with pathways leading to the various effects; plants first, then animals. It consists mainly of reviews; discussion is replaced by a few contributions of later date. On the animal side, diurnal rhythm is well treated, and so is photoperiodicity of insects and of birds; but plumage changes are omitted. One page, on a species said to be unaffected by day-length, supplies the only detail of photoperiodicity in mammals.

JOHN HAMMOND, JR.

Short notices of books

(These notices are descriptive rather than critical and are designed to give a general indication of the nature and scope of the books.)

A Dictionary of Scientific Terms, *by I. F. Henderson and W. D. Henderson (seventh edition, revised by J. H. Kenneth). Pp. xv+595. Oliver and Boyd, Edinburgh.* 1960. *32s. net.*

The dictionary is restricted to the fields of biology, botany, zoology, anatomy, cytology, genetics, embryology, and physiology, in which it gives the pronunciation, derivation, and definition of terms. This new edition contains some 15 600 terms, 1750 new ones having been added. Special attention is paid to American usage, and numerous cross-references are included.

Molecular Distillation, *by G. Burrows. Pp. ix+214. Clarendon Press, Oxford; Oxford University Press, London.* 1960. *35s. net.*

The aim of this book is to bring together the different aspects of the theory and practice of molecular distillation so as to be valuable to the chemist, the physicist, and the engineer. The author considers that his treatment of mixed molecular and laminar gas flow through tubes of arbitrary length is new, and he includes some previously unpublished molecular-distillation composition curves. The special features of gas and vapour flow at low pressures receive detailed treatment.

Zeitschrift für Chemie, *Vol. I, No. I, chief editor G. Horst. Pp. 32. Deutscher Verlag für Grundstoffindustrie, Leipzig. DM 36 p.a.*

This monthly journal provides reviews on various aspects of chemistry and industry, with a view to making it possible for those working in these fields to keep abreast of progress. This issue includes articles on the mechanism and kinetics of surface catalysis, by S. von Roginski, and on the problem of evolution in terms of the thermodynamics of irreversible processes, by I. Prigogine, as well as a number of short original contributions.

Optical Crystallography *(third edition), by E. E. Wallstrom. Pp. x+356. John Wiley & Sons Inc., New York and London.* 1960. *68s. net.*

This book is an introduction to optical crystallography, and is intended for geologists, mineralogists, chemists, and ceramists. It deals particularly with the theory and practice of the use of the polarizing microscope, avoiding the more difficult and abstruse aspects, and also using drawings to avoid mathematics. The third edition has a completely new text, and has an added chapter on crystal-rotation methods.

Tools of Biological Research, *second series, edited by H. J. B. Atkins. Pp. xii+175. Blackwell Scientific Publications Ltd, Oxford.* 1960. *37s. 6d. net.*

The techniques described in this book, the account of a symposium held at Guy's Hospital in 1959, include electron-spin resonance spectroscopy, paper chromatography, the ultracentrifuge, fluorescent-antibody techniques, phonocardiography, and vectorcardiography. It is mainly directed to surgeons engaged in research, but it contains material that will also interest other research workers in applied biology.

Ultrafiltration, *by L. Ambard and S. Troutmann. Pp. x+67. Charles C. Thomas, Illinois.* 1960. $4.50.

The process of ultrafiltration is interesting because some aspects of it seem analogous to processes that occur at the boundaries of living cells. This monograph deals first with the general principles of ultrafiltration, basing the explanations on a consideration of changes of hydration of the ions involved. The discussion is continued to include the ultrafiltration of plasma and albumin solutions, and the biological implications of these studies.

Introduction to Entomology, *by R. Jeannel. Pp. 344. Hutchinson & Co. Ltd, London.* 1960. *63s. net.*

The information in this book is arranged under three general headings, anatomy and classification, biology, and paleontology and geographical distribution. The individual chapters cover these fields in some detail, giving particular attention to classification, to the evolution of insects, and to fossil species.

Physiology of Plants, *by P. Font Quer. Pp. 128. Arrow Books Ltd, London.* 1960. *Paper back, 5s. net; bound, 10s. 6d. net.*

This is one of the Arrow Science series; it is a translation, from the Spanish, of part II of *Botánica Pintoresca.* It is written for the educated layman and for those starting the study of botany, and it covers such subjects as the functions of water, the chemistry of plants, genetics, and plant movement.

Industrial Electrical Furnaces and Appliances *(second edition), by V. Poschkis and John Persson. Pp. xvi+607. Interscience Publishers Inc., New York; Interscience Publishers Ltd, London.* 1960. *£9 net.*

The intention of this book, originally published in two volumes, is to place the practice of electrical heating on a rational basis and eliminate, as far as possible, designing by empirical methods. Heat transfer is studied particularly carefully, as the authors consider that engineers are not always sufficiently appreciative of its relevance to furnace design. The new edition also includes information on oxygen lancing and induction stirring, and on special furnaces, such as consumable-electrode furnaces and skull furnaces.

Hospital Infection: Causes and Prevention, *by R. E. O. Williams, R. Blowers, L. P. Garrod, and R. A. Shooter. Pp. x+307. Lloyd-Luke (Medical Books) Ltd, London.* 1960. *35s. net.*

This is essentially a practical book, intended for surgeons and sister tutors as well as for bacteriologists and pathologists. The first part of the book deals with the epidemiology of hospital infection, including chapters on staphylococcal infection, gastro-intestinal diseases, and tetanus and gangrene. The second part deals with the control of hospital infection, and considers all aspects, including the design of operating theatres, the use of antibiotics, and the different methods of sterilization.

Notes on contributors

F. G. A. STONE,

M.A., Ph.D.

Was born in Exeter, England, in 1925 and was educated at Exeter School and Christ's College, Cambridge. From 1951 to 1953 he was a postdoctoral research fellow at the University of Southern California, Los Angeles. In 1954 he became a member of the staff of the Department of Chemistry at Harvard University, where he is engaged in teaching and research in the field of inorganic chemistry. He has published over sixty scientific papers and is a contributor to several books.

SIR HARRIE MASSEY,

B.A., M.Sc., Ph.D., F.R.S.

Was born in 1908 and was educated at the University High School, Melbourne, and Melbourne University. After graduation he undertook research work in atomic physics at the Cavendish Laboratory, Cambridge. In 1933 he was appointed head of the department of mathematical physics at Queen's University, Belfast, and in 1938 became Goldsmid professor of mathematics at University College, London. He held this chair until 1950, when he became Quain professor and head of the department of physics at the same college. He is chairman of the British National Committee for Space Research. In addition to many papers and reviews in scientific journals he is joint author of two monographs, 'The Theory of Atomic Collisions' (with N. F. Mott) and 'Electronic and Ionic Impact Phenomena' (with E. H. S. Burhop); he has written five other books, 'Ancillary Mathematics' (with H. Kestleman), 'The Upper Atmosphere' (with R. L. F. Boyd), 'Negative Ions', 'Atoms and Energy' and 'The New Age in Physics'.

R. C. NAIRN,

M.D., Ph.D.

Was born in Liverpool in 1919 and was educated at Liverpool Institute and the University of Liverpool Medical School, where he graduated in 1942. In 1947 he was appointed Lecturer in Pathology, University of Liverpool, where he carried out research into the pathogenesis of oedema for the next four years. In 1952 he moved to his present appointment of Senior Lecturer in Pathology, University of Aberdeen, and Honorary Consultant Pathologist to North Eastern Region (Scotland). He continued investigative work, at first mainly concerned with the relationship of kidney to body water and blood pressure, in 1955 during tenure of a Bunt's Research Fellowship at the Cleveland Clinic, Ohio. Since then his attention has been devoted to immunopathology and the development and application of the fluorescent tracing method.

G. LEWIN,

Ph.D.

Was born in 1907 in Berlin, and received his education at the Universities of Berlin, Munich, and Prague. He was associated with several industrial research laboratories; his main activities were in the studies of electron physics and gas discharges. In 1957 he joined the Plasma Physics Laboratory of Princeton University, in Princeton, New Jersey, and at the present time he is the head of the vacuum group.

E. P. ABRAHAM,

M.A., D.Phil., F.R.S.

Was born in 1913 and was educated at Queen's College, Oxford. He is Reader in Chemical Pathology at Oxford University and a Fellow of Lincoln College. He was a Rockefeller Foundation Travelling Fellow at the University of Stockholm in 1939 and the University of California in 1948. During the war he was associated with the work carried out on penicillin at Oxford. Since 1949 he has studied the chemistry and biochemistry of a number of peptide antibiotics and other natural products with biological activity. In 1957 he gave the Ciba Lecture in Microbial Biochemistry at Rutgers University, New Jersey, and in 1960 was a guest lecturer at the University of Sydney.

G. G. F. NEWTON,

M.C., M.A., D.Phil.

Was born in 1919 and graduated at Cambridge University in 1947. He is at present at the Sir William Dunn School of Pathology, Oxford, which he joined in 1947. He received his D.Phil. degree in 1950 and since 1953 has been a member of the external staff of the Medical Research Council. His work has been mainly concerned with the isolation and properties of naturally occurring peptides with biological activity.

J. D. PYE,

B.Sc.

Was born at Mansfield, Nottinghamshire in 1932, and was educated at Queen Elizabeth's Grammar School, Mansfield; he obtained a degree in zoology at University College of Wales, Aberystwyth, in 1955, and started research there into the nervous control of colour change in teleost fishes, and later continued this at Bedford College, London. He is now a research physiologist at the Institute of Laryngology and Otology, University of London.

Investing in research

The pursuit of knowledge for its own sake has long been represented as the highest form of human endeavour, and this is certainly a proposition with which few scientists will quarrel. Moreover, it can be argued that scientific instinct and the public interest go hand in hand, for there is no lack of examples of discoveries made without thought of application that have in due course proved of the greatest practical importance. On the evidence of past experience it is undeniable that, in the long run, scientific research is bound to produce useful results, even though their nature may be unpredictable. This is a comforting situation, for it may be held to be sufficient justification for almost any sort of inquiry. Nevertheless, the total expenditure on research is now so great that those concerned with its allocation must increasingly consider the quantitative as well as the qualitative side of the argument. Exact estimates of research expenditure are impossible to make, because one can neither define research precisely nor follow its ultimate ramifications. Approximate calculations are feasible, however, and a recent estimate by the Advisory Council on Scientific Policy shows that British expenditure in 1959 was £478 million, representing 2·35 per cent of the gross national product. Of this total, rather more than half represented research and development in private industry and almost exactly one-third was government expenditure, mainly on defence; university research represented rather less than five per cent of the total. Although the United States total, calculated on the same basis, is many times greater (£3971 million), the percentage of the gross national product is much the same, namely 2·74 per cent. Strict comparisons with other countries are not possible, but there is good reason to suppose that expenditure in Russia is relatively at least as high, if not higher, and most of the other leading nations also devote very large sums to scientific research. Nevertheless, it is still generally acknowledged that still more must be spent if we are to make the best use of scientific possibilities. How great an advantage the further vast expenditure contemplated will be depends to a great extent upon its allocation: if well spent, the greater total may, by virtue of results, represent a smaller percentage of the gross product.

The present world expenditure on research and development is enormous, both absolutely and if measured against the comparable figures for the years before the first world war, when so many still-current ideas about policy were formed. At that time science was generally far more remote from everyday life than it is today, and if the research worker had to justify himself at all it was only to those who supported him. Today, however, the situation is very different, for much of the money comes directly or indirectly from public sources. In the long run it is essential, if support is to be continued and increased, to convince the policy-forming section of the public that it is receiving value for money. Rightly or wrongly, in this respect value for money means sufficient practical application—considered in the widest sense of a higher standard of living, better health, defence, national prestige, and so on—to justify the overall outlay. A substantial part of research must now be financed from the wealth that its application creates, for the total sum involved has become far too great to be spent without thought of practical advantage. This is the hard reality of the situation, unpalatable though it may be to those who believe that scientists should be the only arbiters of things scientific. If science is not merely to flourish on its present scale but to expand, public support and understanding of its aims are absolutely essential.

This does not, of course, at all imply that we must do more and more applied research to the detriment of fundamental work, for, apart from considerations of scholarship, that would ultimately be ruinous. But it does mean that, as a matter of common prudence, more thought must be given to the factors that determine where the balance between these two fields of research should properly be struck. That there is a point of balance is evident. There are types of investigation that, so far as can be seen at the time, promise no useful applications whatever. In a balanced community such pursuits deserve support for the same sort of reasons as lead to encouragement of music, drama, literature, art, and other aesthetic interests; they are part of the general cultural background. At the other end of the scale, even the most convinced materialist can see that a term is set to the useful application of knowledge if we do not ensure that our store of it is constantly increased. But the public purse is not inexhaustible, and this in itself means that there can be only limited support for all kinds of research. Where should the balance lie?

Certainly no hard and fast answer can be given to this question, for, apart from other considerations, conditions vary greatly between the different branches of science. In some, progress is delayed for lack of fundamental knowledge; in others, knowledge is accumulating faster than it can be turned to practical account, and we have, as it were, capital earning no interest. Today, the different branches of science increasingly cross-fertilize each other, and it might seem to be logical to increase expenditure in the weakest fields so that progress may take place fairly evenly on a broad front. But this, if it were feasible at all, could be done only on a long-term basis, for merely making money available is only the beginning. Not only must laboratories and equipment be provided, but, far more difficult, the trained men must be found to use them. In science, as in other pursuits, economic pressures are heavy, and incentives can be given to encourage workers to enter fields where they are particularly needed. But this, again, is by no means the whole answer; what really produces results, in science as in so many other branches of co-operative endeavour, is enthusiastic leaders, and of these the supply is very limited.

But neither leaders nor led create themselves, and the education of the potential scientist is of the first importance in maintaining and increasing the number of active workers. On a long-term view this is something that may properly be considered as part of the cost of research, though it is too rarely looked upon as such; nevertheless, it is indisputable that the quality of a scientist owes much to his early training. At the moment, virtually every country in the world is making great and costly efforts to multiply its output of science graduates; yet illogically small attention is devoted to ensuring adequate teaching at school level, during the impressionable years in which the principles of the scientific method are most firmly inculcated. In most countries, little has been done to make the reward of those concerned with preparing the next generation of scientists comparable with that of working with the present one. Here, too, there is clearly a lack of balance, though it is easier to diagnose the fault than to suggest a remedy. Nevertheless, the mere realization that we are letting the edifice grow without paying sufficient attention to its foundations is an important step towards a reappraisal of the apportioning of scientific expenditure.

The need to strike a proper balance between pure and applied research has already been touched on, and it presents a problem of the greatest importance for the future of science. The belief dies hard that the former is in some way intellectually superior to the latter; during their training the best men tend to be imbued with the idea that the only worth-while goal is to remain in academic research, industry being represented as the refuge of the second-rate. Though the belief today is held only by a minority, this is sufficient to add to the difficulty of recruiting good men for applied research. This position would be justifiable if the basic premise were correct, for it would be wrong to use the best brains at anything less than their full capacity. Yet all the evidence is to the contrary; those who have experience of both have no doubt that for the very great majority of scientists research in industry can offer problems no less intellectually challenging than are encountered in academic laboratories. This does not mean, of course, that all industrial science is exciting, for necessarily much of it is of a routine nature. But much the same may be said of a good deal of academic research. The collecting of essential information may well be mundane and repetitive; it is in devising means of gaining new knowledge, and in the interpreting of observations, that the original mind can best exercise itself. In considering the general question of expenditure on research it is necessary to think not only how the total should be apportioned between pure and applied work but what steps need be taken to ensure that both get a proper share of the best workers. Until this is done, the full advantages of the application of new knowledge will be difficult to realize. Moreover, if the present prejudice lingers, the purists themselves will suffer, for it is undeniable that the wealth created by applied science makes a considerable contribution to sustaining pure science; in the long run, the fortunes of both are inextricably linked.

Gravitational waves

H. BONDI

The experimental study of gravitational waves poses great problems, chiefly because of the extreme weakness of gravitational fields; the waves could become appreciable only in the case of very massive bodies moving with high accelerations. Although there is no present prospect of our being able to observe gravitational waves, some of their principal properties can be inferred. This article shows how gravitational theory is related to the rest of physics, and reviews some of the lines along which it has developed in recent years.

Gravitational waves are not only unfamiliar, even by name, but are distinctly unlikely to be observed. Despite this, their study has received appreciable attention, and is instructive because it is linked to so many branches of science. It stands between the physics of large-scale phenomena and quantum theory, between gravitational theory and the subject of electromagnetic and other waves, between pure mathematics and physics. Present investigations are concerned with the gravitational fields of rapidly moving bodies. Gravitation is a negligible force unless the body exerting it is very massive, and high accelerations are virtually unknown among such bodies: hence the experimental intractability of the subject. One is necessarily confined to mathematical investigations, based on the best available theory of gravitation, namely Einstein's general theory of relativity; many of the difficulties of the subject are due to the mathematical complexities of this theory. To make clear why enough confidence is felt in this theory to justify the effort spent on gravitational waves, it is necessary to go far back into physics and consider the basic character of the theory. Because our subject is interwoven with so many other branches of physics, it will be necessary to review certain aspects of these in some detail before the problems of gravitational waves can properly be formulated.

NEWTONIAN DYNAMICS

Though it is well known that Newton's three laws put dynamics on a firm footing and enabled it to enjoy a period of rapid growth, the exact nature of his contribution is not always clearly understood. It will be recalled that Newton related force to acceleration, whereas previously it was believed to be related to velocity. Why was it such an important contribution to make what was effectively a new definition of force, in which acceleration replaced velocity? The reason is clearly seen when the motion of the Earth is considered. If one considers the force responsible for this motion and presumes it to be related to its velocity, as in pre-Newtonian days, then one should be looking in the direction of the Earth's velocity to find something responsible for its motion. But looking in the direction of the Earth's velocity is fruitless; sometimes one sees one rather insignificant-looking star, then another, and at other times virtually nothing. If, however, one looks in the direction of the Earth's acceleration one sees the Sun, and it is very acceptable indeed to regard this most important body as being in some way the cause of the orbit of the Earth. To put it a little differently, Newton's definition of force enables one to link a kinematical observable, the acceleration of the Earth, with something that can be observed optically, the direction of the Sun. It is a characteristic of any successful scientific hypothesis that it links two previously unrelated kinds of observation, and this is achieved here; it is in this way that Newton's dynamics proved so eminently successful. Moreover, we observe here a point that will recur several times in this article, namely the essential unity of physics. Physics cannot be subdivided into a number of watertight compartments, such as dynamics, optics, and heat; these fields are all linked with each other in numerous ways, and many of the most important physical insights would be lost if the links were not borne in mind. We cannot have one type of physics for dynamics and another type of physics for another subject; dynamics could never have been established without optical observations, and there is no such thing as a purely dynamical experiment. More than one branch of physics is always concerned, and, as we have just noted, the very foundation of dynamics depends on the union of a dynamical and an optical observation.

The next point to which attention must be directed is the nature of the system of reference relative to which acceleration is measured. Force is regarded in Newtonian dynamics as something essentially real, something of true physical significance. Newton's second law implies, therefore, that acceleration is similarly real. But acceleration is the rate of change of velocity, and velocity is the rate of change of position, and position, velocity, and acceleration are all in some way relative. The state of motion of the measuring system decides what values one obtains for each of these variables. Thus, if acceleration is something physically real, something that can be identified with a true physical variable, force, then only those systems of reference that give acceleration the correct value are admissible for our observations. If we have such a system of reference and take a second system of reference in unaccelerated motion relative to it—that is, moving with constant velocity in a fixed direction —then the velocities measured in the second system will differ from the velocities measured in the first system, but the accelerations measured in the second system will be the same as those measured in the first system. The second system will therefore be just as admissible as a standard reference system as the first one. Therefore, in Newtonian dynamics we do not have a unique system of reference, we do not have some kind of standard of rest for the Universe at large, but we have a whole family of such systems of reference, any two members being in relative unaccelerated motion. Any such system of reference is called an inertial system, and acceleration as observed from an inertial system is connected with force through Newton's second law. There is therefore a very sharp distinction in Newtonian dynamics between two kinds of frames of reference, namely inertial frames and non-inertial; only the forces measured in inertial frames are regarded as real. If we work in an accelerated frame of reference, that is, a non-inertial one, then fictitious forces arise. The best known of these is the centrifugal force, which arises when we work in a rotating system of coordinates, and there are several others of that kind.

Since all forces are found to be the same, from whichever inertial system they are viewed, it follows that all inertial systems are dynamically completely equivalent. One can therefore state the Newtonian principle of relativity in the following way:

1. There exists a sharp distinction between inertial systems and non-inertial systems.
2. Any two inertial systems are in relative unaccelerated motion.
3. Any two inertial systems are dynamically completely equivalent, and it is impossible by any criterion of internal dynamical experiments to distinguish one from another.

The significance of this last principle is simply that if, for example, we are in a smoothly moving aircraft and conduct experiments within this aircraft, then all these experiments will give precisely the same result as such experiments would give in a room on the ground. As there are no physical effects of velocity, but only effects of acceleration, it follows that one cannot discover the velocity of any object by internal experiments. Velocity is not absolute but purely relative, whereas acceleration is absolute.

In addition to these basic principles, simple mathematical rules of transformation can be set up that relate measurements made in one system to measurements made in another, whenever this is required; that is, when objects external to both these systems are being measured. In these transformations, acceleration is naturally preserved, and also time, which in Newtonian physics is considered as a universal and absolute quantity.

ELECTROMAGNETISM AND SPECIAL RELATIVITY

In Newtonian dynamics, velocity is never regarded as something fundamental; indeed, it is, in a certain sense, arbitrary. In any field of force, in any conceptual physical situation, we can start off at a given moment by assigning arbitrary positions and velocities to all the particles. The laws of dynamics then specify their subsequent behaviour. In other branches of physics, however, and especially in electromagnetism and the associated field of optics, there exists a velocity of fundamental importance, namely the velocity of light. Not only is this the velocity in a vacuum of all the many kinds of electromagnetic waves, but it also occurs in the basic equations of electromagnetic induction. If the spirit of the principles of Newtonian relativity is applied, it follows that this basic feature, the velocity of light, should be independent of the inertial system from which it is observed. The assertion that this is so, and, therefore, the re-affirmation of the Newtonian principles of relativity for the electromagnetic case, is known as the special theory of relativity. Stress is so often laid on the revolutionary character of this theory that it is frequently overlooked that its re-affirmation that Newtonian ideas are just as applicable to electromagnetism as to dynamics is simply an assertion of the validity of Newtonian relativity for the whole of physics. Thus the theory is, in a sense, profoundly conservative; the principles that have been found to hold for dynamics are stated to hold for the whole of physics. The special theory of relativity is a necessary consequence of any assertion that the unity of physics is essential, for it would be intolerable for all inertial systems to be equivalent from a dynamical point of view yet distinguishable by optical measurements. It now seems almost incredible that the possibility of such a discrimination was taken for granted in the nineteenth century, but at the time it was not easy to see which was more important—the universal validity of the Newtonian principle of relativity or the absolute nature of time, referred to previously.

A simple piece of mathematics shows that the velocity of light cannot be the same in all inertial frames of reference if the assumption of a universal time is maintained. One has to drop the metaphysical idea that time is something absolute, and regard time in the proper physical sense as that which is measured by a clock. In modern science it is becoming more and more understood that it is pointless to talk about quantities unless it is stated how they are to be measured. It is not scientific to regard uniform even-flowing time as being in some mysterious way unalterable. Nor can we assume that, in some ill-defined sense, the time measured by a clock in one system of reference equals the time measured in another system. What we can assert by taking over the Newtonian principle of relativity, which in its application to the whole of physics is called Einstein's special principle of relativity, is the equivalence of all inertial systems. New transformation rules have to be stated, and their essential quality is that time becomes linked with space. This is generally referred to as the four-dimensional nature of space-time. Although this phrase is sometimes regarded as mysterious, the basic idea is perfectly straightforward. What we mean by the three-dimensional nature of space is that we require three different measurements to specify the extension of a body, but that by a rotation they may become mixed up with each other. They are not in any sense absolute, although we may work out from them quantities, such as the distance between two points within the body, that are independent of the way in which we look at it. We mean by the four-dimensional nature of space-time that the velocity of the observer may lead to a mixing up of time-interval with the space-extent of an object in just this manner. The essential point is that this effect is very similar to that of a rotation on the three space-coordinates, and that just as certain measurements, such as the distance between two points of the body, are independent of the aspect, so certain other types of measurement, referred to as the interval, are independent of the state of motion of the observer.

It is not necessary here to refer to the innumerable experimental tests of the theory of special relativity or to its many well-known achievements, such as the identification of mass and energy. It will, however, be of advantage to consider its algebraic power, which arises directly from the four-dimensional nature of space-time. There are few simple algebraic structures that fit into space-time, and this fact enables very important deductions to be made. We recognize that in ordinary space of three dimensions the possibility of a change of position of the observer reduces the number of physical laws that are permissible. It is permissible to say that the force on a moving charged particle due to a magnetic field is at right angles to the velocity of the particle and to the direction of the magnetic field. But to say that the force is always in a direction to the right of the particle is not permissible, because this would mean different things to different observers.

Let us now imagine a stage in the progress of science when the laws of propagation of light were well known, enabling the special theory of relativity to be established, but when the laws of electromagnetism were still very little known and the electromagnetic nature of light was not understood at all. Suppose, in particular, that magnetism were wholly unknown and that only the following two simple laws of electrostatics had been discovered. Firstly, that the lines of electric force originate only at electric charges and that this is true whether the charges are moving or not. Secondly, that the electric field is conservative—meaning that if you take a charge round a closed circuit the total work you can get out of the electric field vanishes—but only if the field is constant in time. How would the situation look from a relativistic point of view? In relativity, we have a space of four dimensions, and therefore every directed quantity (vector) has to have four components that can, to some extent, be transformed into each other by changing the velocity of the observer. Thus in dynamics the momenta are the three space-components of such an entity and energy is the time-component. Throughout the application of relativity one usually finds that the relevant time-component is associated with the three space-components of a vector. However, the law concerning charges and the origin of electric lines of force is not of this kind. Not only does it involve just the three space-components of the electric field, but its validity, even when charges are moving, explicitly denies the existence of a time-component. It follows from this that the electric field cannot be such a simple algebraic structure as the momentum and energy of a particle constitute. Something more complicated has therefore to be examined, and the next in order of complication after a vector is a tensor.

A tensor is an entity having, in four dimensions, sixteen components; however, the number of these components may be reduced by symmetry considerations. In a particularly simple form of symmetry (antisymmetry) the number of components is reduced to six, three of which form one group and three another, each being transformed into the other by motion. The three components of the first group can readily be identified with the three space-components of the electric field and mathematically cannot have a time-component associated with them. The three relevant components of an antisymmetric tensor are the only simply occurring set with this property. Associated with them are the three other components, and as these arise from velocity, our relativist could deduce that, just as electric charges at rest produce an electric field, electric charges in motion (an electric current) should produce another kind of field. This is what we call the magnetic field, and the relativist would have deduced its existence from simple considerations of symmetry. Moreover, the fact that the electric field in its static condition was conservative will have enabled him to go a step further in identifying the structure of the electromagnetic field, for he would have been able to deduce that the entire field can be derived from a potential vector in four dimensions. Then, linking the field, the entire complexity of which he could thus have deduced from relativity considerations, with the sources of the field (the charges at rest and in motion), he would have arrived at the full set of Maxwell's equations describing the electromagnetic field. He could then have deduced from these equations that changes in the electromagnetic field would be radiated and propagated with the velocity of light. To complete his set of equations he would naturally have to make an assumption about causality, that is, that news of changes in a system travels forwards in time and not backwards. It would not have been difficult for him then to identify light with one form of these electromagnetic disturbances, whose propagation properties he had deduced entirely from relativity and the most elementary rules concerning the electrostatic field.

It may at this point be worth discussing briefly what the mathematician means by waves. He does not necessarily mean a periodic sinusoidal phenomenon; all he means is something that propagates. As such, one of its essential features is its ability to convey information. It is by means of a wave that the receiver obtains information about what goes on at the transmitter. Changes in the state of the transmitter are propagated as waves; the idea of news is inextricably interwoven with the idea of waves, and this is essential for their understanding.

A second characteristic feature of waves is the transfer of energy. There are many forms of energy transfer, and it may be useful if we think first about the way energy is transferred from a power station to the consumer along the supply wires. The important point here is that the energy put out by the power station is entirely determined by what the consumer actually uses, apart from losses in transmission that are small and could in principle be made negligible. In wave propagation, however, there is a wholly different situation. A radio transmitter radiates power into space, and the amount of power put out is independent of the number and location of the receivers. The amount of energy radiated is quite independent of whether people switch their receivers on or not. This is an amazing property of radiation which in many ways is difficult to understand, for what happens to the energy radiated? It seems to be swallowed by space at large distances. There are good reasons for supposing that the system works only because all radiation is eventually absorbed, though at very great distances, and moreover that this is closely connected with the actual construction of our universe, which is expanding. In the present context the two crucial characteristics of waves may therefore be stated to be that energy is radiated irrespective of the existence and location of receivers, and that waves can convey new information.

GRAVITATION

Gravitation is distinguished from all other forces by the fact, discovered by Galileo, that all bodies fall equally fast. The force of gravitation is proportional to the mass of the body, so that the acceleration of a body is independent of its mass. The full significance of this remarkable phenomenon will be discussed a little later, but when Newton formulated his theory of gravitation he simply took it as axiomatic that the pull of gravitation on any body was proportional to its mass. Furthermore, he assumed the inverse square law of force and that gravitation was a conservative force. Finally, he assumed that the gravitation-producing property of matter, the source strength of the gravitational field, also derived from its mass, a result that in any case follows from his third law concerning action and reaction. A tremendous amount of work has been done on the basis of this Newtonian formulation of gravitational theory. The solar system has given vast scope for applying Newton's theory of gravitation and for checking it over long periods of time in the minutest detail.

Sixty years ago it could be said that Newton's theory of gravitation was the best established and soundest in physical science. Then, quite suddenly, a new theory of gravitation was advanced, Einstein's general theory of relativity, and belief in the absolute validity of Newton's theory was abandoned with little resistance. How could it come about that such a well-established theory was dropped and replaced by another one whose chief merit is that it gives the same results as Newton's theory? The very few and minor points of discrepancy are observationally not too firmly established, and yet the conceptual basis of Newton's theory is regarded as stone dead and general relativity is the theory on which present work is based.

There are two reasons for this eclipse of the Newtonian picture, of which the first is simply special relativity. It was shown in the last section how the basic notions of optics and electromagnetic theory demand that we abandon the Newtonian idea of a uniform even-flowing time and a three-dimensional space and substitute for them a four-dimensional space-time. Newtonian gravitation does not fit into this four-dimensional space-time; it does not have the correct algebraic structure. That this had not become apparent earlier is simply due to the circumstance that all our knowledge of gravitation arises from rather slowly moving bodies—and for such bodies there is almost no difference between Newtonian space and time on the one hand and the space-time of special relativity on the other. Nevertheless, the complete acceptance of special relativity implies that Newtonian theory cannot be fitted into the framework of modern physics and must therefore be abandoned. The choice is now severely limited, as the number of simple relativistic theories is small, and the first suspicion of the existence of gravitational waves arises at this stage. We have seen how the simple laws of electrostatics, including the inverse square law, lead virtually inescapably, within the framework of the space-time of special relativity, to the notion of electromagnetic waves. As soon as one decides that gravitational theory, too, must be relativistic, one suspects that the analogous development should take place here; that is, that while static gravitational fields are well described by the inverse square law, the gravitational fields of fast-moving bodies should give rise to gravitational waves.

The second reason for rejecting Newtonian theory—the principle of Galileo—serves to determine which of the several possible relativistic theories of gravitation should be accepted. In Newtonian theory, the fact that gravitational force is proportional to the mass of the body on which it is acting is considered no more significant than, say, that electrostatic forces should be proportional to the charge of the body on which they are acting. However, this view of the association of mass with gravitational force is rather shallow. Bodies can have any charge we please to give them, but the mass of a body is inalienable. There is no such thing as a body without mass, though there are plenty of bodies without charge. This very directly affects Newton's first law, the very basis of Newtonian dynamics, which states that a body on which no forces are acting moves in a straight line with constant velocity. For what are the chances of testing this law? Suppose we wish to experiment by observing the motion of a suitable body. In order to test the first law we have to remove all the forces acting on it. First, therefore, the physicist carries out an inspection to make sure there are no ropes pulling the body, then he checks that there is no friction between it and its surroundings. Next, he makes certain that it is without charge, so that it is not pulled by electric fields, and so on. In this way he can remove all the forces acting on the body, or rather all the forces with one exception—gravitation. He cannot relieve the body of its mass, because this can no more be separated from it than a man from his shadow.

Newton thought he could get round this difficulty by specifying that the body to be tested should be sufficiently far from any other bodies, but this would scarcely serve the purpose. Not only should the standard motion, which is what the first law prescribes, be available at every point, but it is now known that the universe is everywhere populated with matter, so that no sufficiently isolated region exists. We have, therefore, to conclude that the Newtonian standard is not attainable. The basic premise of the Newtonian standard is that if a body follows the standard motion, then there is no purpose in looking for a force causing it. It is only when the motion of the body deviates from the standard that one has to look for a force. However, this sort of argument fails if the Newtonian standard is not an attainable, testable standard. The only way out is to take as the standard motion of matter its motion under gravitation; in other words, to abandon the whole notion of gravitation as a force. A force is something that causes matter to deviate from its standard motion, the standard being its motion under gravitation and inertia. Gravitation and inertia now become, in some sense, one; they provide the rails on which matter runs when there are no forces present. Naturally, in order to include all the complicated phenomena of gravitation, the new standard motion must be infinitely more complex than motion in a straight line under constant velocity, the Newtonian standard. Nevertheless, a far more satisfactory situation has now been reached, in which due account is taken of the fact that gravitation makes all bodies move equally, and in which the new standard is available.

What remains of the gravitational field when we regard it as the standard? The only property that we can work with now is its inhomogeneity, the fact that neighbouring particles do not fall in precisely the same way, because the gravitational field varies from point to point. A study of the relative acceleration of neighbouring particles now serves to describe the gravitational field, and this can readily be done in a relativistic fashion. Moreover, the form of the relativistic theory of gravitation has now been fixed; it must be such that gravitation reproduces the effects that inertia can produce, because inertia and gravitation are now regarded as one. It will be recalled that in Newtonian theory, when a non-inertial system of reference was used, fictitious forces such as centrifugal force arose. It follows that the description of the gravitational field must correspond to the complexity of motions that can be produced by using arbitrary systems of reference and considering the fictitious inertial forces then arising. On this basis it turns out that gravitation is described by a symmetric tensor in space-time, that is, an entity with ten independent components. This tensor has in fact the special significance that it describes the nature of space-time itself. Space is naturally measured out by using freely moving particles and light rays, the latter being a limiting form of freely moving particles. Therefore, the new standard motion serves to describe space-time itself. It is measured in terms not of straight lines traced with constant velocity, but in terms of the free motion of matter.

This is the foundation of Einstein's general theory of relativity, and it can be seen how immensely compelling it is. Its basis is that all bodies fall equally fast, that being affected by gravitation is an inalienable property of matter. It is, from the start, conceived in space-time, and therefore it fits the notions of electromagnetism and special relativity; consequently it is the natural successor to Newton's theory. There the equation that determines the way in which matter produces the gravitational field consists of a certain linear combination of relative accelerations of neighbouring particles, the 'Laplacian', being equated to the density of matter. A similar procedure in general relativity yields the so-called field equations, which again use linear combinations of the coefficients determining the relative accelerations of neighbouring particles, and equate these to the material content of space. Although this procedure is perhaps not unique, the argument is so compelling that one has little doubt that general relativity is by far the best relativistic formulation of gravitation that also takes account of the fact that all bodies fall equally fast. Moreover, it is shown without great difficulty that, for slowly moving bodies, general relativity gives almost the same answer as Newtonian theory; the discrepancies are very small. One of them concerns an extremely small effect on the motion of the planet Mercury, where Newtonian theory did not quite fit the facts, and there general relativity does. Two other discrepancies are concerned with the propagation of light. One of them can be inferred from very basic principles but is difficult to observe experimentally, though this has been achieved quite recently. The other concerns the effect of gravitation on light rays, and here again the observational results are in fair agreement with relativity, which demands only a very small effect.

To recapitulate, we are forced to abandon

Newtonian gravitation because it is not relativistic. If we try to formulate a relativistic theory of gravitation we are strongly impressed by the need to take account of the fact that all bodies fall equally fast, and then one is driven, almost inescapably, to Einstein's theory. Experimentally it goes beyond Newton's theory, though not very much so. But the burden of the argument here has been that it is the compulsion coming from other branches of physics, more than the agreement with those details of observation where there is a difference between Newton and Einstein, that forces one to adopt general relativity. We are driven to it by the principle of the unity of physics that demands that electromagnetism and optics should not exist in a different sort of space-time from that of gravitation.

THE QUANTUM THEORY

Only a little need be said here about the important understanding that has been gained in the last fifty years of the structure of atoms and nuclei, and of the behaviour of sub-atomic particles. It is well known that the physics governing these differs from that for larger bodies, because, fundamentally, the effect of the means of observation on what is observed can no longer be neglected. When dealing with the smallest entities, one cannot use tools finer than the entities themselves. This introduces indeterminacies and uncertainties that are formalized in quantum theory. The first step here was the quantization of dynamics (quantum mechanics), and the first application was to the motion of electrons in the field of a charged nucleus. This field is electromagnetic but it is effectively only electrostatic. By a fundamental property of quantum theory the electrostatic field itself need not be quantized, essentially because it cannot convey news; it is treated simply as an aura surrounding the central nucleus. At this early stage of the development of the theory it was therefore only in the response of bodies to known forces that allowance had to be taken of the indeterminacies arising from the restrictions on observability. Quantum mechanics in the strictest sense is sufficient for the discussion of the basic features of atomic structures, because there the field is of the special electrostatic type, obeying the inverse square law, that does not need to be quantized. This step, quantization, was only later taken for the electromagnetic radiation field, where it led to grave complications and difficulties that have not been overcome until quite recently.

GRAVITATIONAL WAVES

It is a fundamental demand of quantum theory that there should, beyond a certain degree of accuracy, be no way of ascertaining simultaneously both the position and the momentum of a particle. The gravitational field due to a body is of appreciable strength only if its mass is quite large. On the other hand, the gravitational effect is additive, that is, one half of a mass produces one half of the gravitational effect of the whole mass. If this rule is taken to its limit—and there exists no evidence that it may not be—then there should even be a gravitational field due to a single sub-atomic particle, such as a proton, or neutron, or even electron. Of course, these gravitational fields are incredibly weak, because the masses of the particles are so small, and they are completely hidden by the electrical fields which are immensely stronger, the factor being 10^{39}. Nevertheless, if it is in principle impossible to measure simultaneously both the position and the momentum of the particle, then it must be impossible to find these quantities even by an unbelievably exact determination of the gravitational field. The gravitational field, too, must therefore be quantized to accord with the rest of physics. Once again it is the unity of physics that makes a demand, although so far as actual measurement is concerned we are very far indeed from being able to test the need for it. Logically, however, this does not affect the necessity for quantizing the gravitational field, which must be done to ensure that the same limitations of observability of minute particles apply to gravitational as to other experiments. This difficult process need not be carried out, however, if the gravitational field is simply an aura surrounding the particle. Just as in quantum theory it is not necessary to quantize the inverse-square-law electrostatic field of the central particle, so it would probably not be necessary to quantize the gravitational field if this were purely an inverse-square-law field. If, however, there exists in gravitation a radiation field, then this must be quantized just as the electromagnetic radiation field had to be. Thus we are driven from two different directions to ask whether gravitational waves exist: from the problem of the analogy with electromagnetism, where the application of relativity led from static considerations to wave fields, and from quantum theory, which demands the quantization of the gravitational field if there are such things as gravitational waves.

Finally, there is just ordinary curiosity. We have had a satisfactory theory of the gravitational field, Einstein's theory, for over forty years. It is time we knew what this theory says about the gravitational field of rapidly accelerating particles, even if such a field cannot be observed in nature for what appears to be a quite accidental reason, namely, the absence of sufficiently rapidly accelerated objects of large mass. Thus the demand for unity in physics makes the problem of the existence of these waves interesting, although their experimental verification is unlikely for quite some time.

Having posed the question, one is surprised that it is difficult to get the answer, but the reasons for this are largely mathematical. The equations describing general relativity are, in all but the simplest applications, exceedingly complex and difficult to unravel. Moreover, there is a physical pecularity of their structure. Owing to the unification of inertia and gravitation they allow us to use any system of reference, and it is not easy to specify a system of reference sufficiently without violating the flexibility of space-time that is such an essential feature of the theory. Various methods of approximation have been used, and some of these, at least, are rather suspect. However, considerable progress has been made in recent years and a number of results have been derived.

Firstly, how can gravitational waves be generated? The simplest idea that comes into one's mind concerns the propagation of gravitational changes and could be put into the form of a question. If the Sun suddenly ceased to exist, how soon afterwards would the Earth leave its orbit? However, this is a nonsensical question. The laws of gravitation involve the law of conservation of matter, and so, within the context of gravitational theory, the Sun cannot cease to exist. Next, one might consider what would happen if, instead of ceasing to exist, the Sun suddenly moved off at right angles to the plane of the Earth's orbit. How long would it be before these effects began to show themselves? Again the question is nonsensical, because the laws of gravitation imply the conservation of momentum, and therefore the Sun cannot go off sideways. There are, however, changes that are not forbidden. The Sun could suddenly develop into a prolate or oblate spheroid, and such changes would undoubtedly affect the Earth's orbit. The Sun could explode into two parts going off in opposite directions. These are perfectly reasonable possibilities, and any theory should, at least in principle, clearly state what would then happen.

It turns out that the velocity of propagation of gravitational information is the velocity of light; since this is built into the structure of space-time, the result is in no way surprising. Throughout physics the transport of information is inextricably linked with the transport of energy, though no simple rule exists to determine how much energy is required to convey a given amount of information. To study the question posed above, the relativist considers a body surrounded by empty space and asks how changes in the shape of the body affect its mass. Changes of shape imply changes in the gravitational field, and thus necessarily also imply propagation of information, while the mass of a body is the most convenient measure of its energy. Crudely put, the question is whether changes of shape radiate away any of the mass of the body.

To formulate this question precisely, it is not enough to require the space surrounding the body to be empty, a phrase which in relativity excludes not only matter, but all other fields except gravitational ones. However, if gravitational waves carrying energy exist, such waves might not only be produced by the body and travel outwards from it, but might also be coming inwards from great distances towards the body and so bring energy to it. To exclude such incoming waves, an 'outgoing radiation condition' has to be applied in addition to the demand that space be empty. This condition is closely associated with the application of the principle of causality, namely that the effects of changes of shape should occur only after the changes. For the outgoing radiation condition also prohibits incoming waves specifically arranged to conform to the changes of shape of the body, that might be taken to be waves due to these changes of shape but propagated backwards in time. Thus the condition not only prevents an indiscriminate inflow of energy but a violation of our notions of causality. The formulation of a suitable condition of this type is not easy, and it is gratifying that three different formulations have recently been shown to be equivalent, at least in a case of special importance.

An associated question concerns the origin of the news propagated in the outgoing waves. How can an isolated body come to change shape, especially when it has been isolated for so long that the surrounding gravitational field contains no memory of the body having been put there? Note that the change should in some way be unexpected, so that the waves really convey news. The simplest device would seem to be a time-bomb with a detonator that takes a very long time to go off and is sufficiently defective for its behaviour to be unpredictable. In this way all the conditions of the problem can be satisfied.

The mathematical solution is as yet deplorably incomplete. It has not proved possible to do more than develop some methods of approximation whose validity is not entirely beyond doubt. One of these methods is extremely lengthy and cumbersome, and is so far confined in application to a special case. In another, also of great length, there is some doubt about the unambiguous formulation of the conditions to be satisfied at great distances. A third, less lengthy than the others, confines itself entirely to consideration of the field far from the source. In this last method there is no serious difficulty about satisfying all the conditions at great distances from the source, but one cannot discover which specific distant field is produced by which type of motion of the source.

Nevertheless, some answers have emerged unambiguously, and at least some further questions can be formulated. Thus it is known that if a system that is initially static undergoes some changes of shape and eventually becomes static again, then its mass in the final state is less than its mass in the initial state by an amount proportional to the square of the rate of change of the field, summed through the period of change. The restriction to initially and finally static states is made because only static fields are fully understood and admit of an unambiguous definition of the mass of the system. Thus it has been established that a system can radiate away some of its mass, though this is only a second-order concomitant of changes of shape and not a primary cause of radiation as in the hypothetical case of the Sun's disappearing.

The structure of a gravitational wave must be wholly transverse. Thus, if a wave of limited duration hits a swarm of particles previously at rest relative to each other, then after the passage of the wave there is no relative motion of the particles parallel to the direction of propagation. In one transverse direction the particles will be separating, and in another, at right angles to the first, they will be approaching each other. At large distances from the source a limited section of the wave is almost plane, and plane gravitational waves are well understood. In spite of the complexity of the equations, they travel without change of shape or dispersion. In this, they are like electromagnetic waves (which, however, can be superposed linearly, whereas this cannot be done with gravitational waves) and unlike sound waves, which sharpen up to form shock waves. This constancy of shape is no doubt associated with the transverse nature of the waves.

The construction of a gravitational-wave receiver is also understood, at least in principle. Suppose two massive spheres can slide on a rough stick but are attached at separate positions by springs. Then if a wave travels at right angles to the stick it tends to make the spheres move in relation to one another, as with the particles of the swarm referred to before. This is bound to lead to relative motion between the spheres and the stick, and thus to frictional heating. Some of the energy of the wave has been converted into heat, which is localized energy, and thus the wave has been received. The very response of the receiving system leads to re-radiation, and therefore only a limited amount of energy can be captured by one receiver from a given wave. To make this amount as large as possible, the motion of the spheres must be neither too small nor too large, and thus the concepts of impedance and matching apply in the gravitational as in the electromagnetic case.

A fascinating question arises from the discovery that the distant gravitational field can be changing in a specific, fairly wide, class of ways without there being any radiation. No such change can begin or end without the presence of radiation. What motions of the source can be responsible for such non-radiative changes of the field? Present methods give little indication, but each of two contradictory arguments has some force. According to one, it is lack of acceleration of the particles constituting the source that leads to absence of radiation in spite of source motions. This fits in both with electromagnetic radiation and with the first term of a particular method of approximation, but neither of these arguments is necessarily relevant. The opposing argument stresses that lack of acceleration is irrelevant in gravitational theory because the natural orbits of particles are not the unaccelerated ones, but those of free fall under gravitation. On this view, it is a dust cloud that is likely to be the source of non-radiative changes, because in such a cloud each particle moves freely. But if a dust cloud does not radiate then it does not seem plausible that a passive solid body could; a time-bomb, the prototype of a radiator, is highly active by contrast. Put more mathematically, the relation between pressure and density is time-independent in a dust cloud or in an elastic solid body, but time-dependent in a time-bomb. If there is any form of internal friction—imperfect elasticity, viscosity, and so on—this relation depends on the sense of time, and the processes occurring are irreversible. It is therefore tempting to suggest that the important class of time-dependent but non-radiative fields at large distances are produced entirely by reversible changes in the source of the field, but so far this conjecture is wholly unproved. If it is correct, then there is far less gravitational radiation in the universe than there would otherwise be. The solar system would radiate about one kilowatt in gravitational waves if these were produced by all accelerations, but merely a fraction of a milliwatt if only irreversible processes, such as tidal friction, led to radiation. The question is fascinating, but there does not seem to be any easy way of answering it.

Finally, brief reference may be made to another unsolved problem, that of the applicability of Huygens's principle to gravitational waves. All equations of the wave type prohibit the transmission of any influence at speeds exceeding the characteristic velocity, but only some imply that all influences necessarily travel at just this speed. Thus the ordinary equation of small-amplitude sound-waves shows that in plane and spherical sound-waves the entire influence travels at the speed of sound; for example, after the passage of a brief pulse the medium is wholly at rest again. In the case of cylindrical sound-waves, however, while particles rest in front of the pulse, behind it there is disturbance, because some of the energy travels at speeds lower than the velocity of sound. Thus Huygens's principle, according to which the entire disturbance travels with the characteristic velocity, is satisfied for plane and for spherical sound waves but not for cylindrical ones. In the case of gravitational waves it is known that Huygens's principle applies to plane waves and does not apply to cylindrical ones, but the physically most important case of spherical waves (radiation from an isolated system) is quite obscure, for purely mathematical reasons. A solution of this problem would clear up many other uncertainties in the subject.

We are still quite some way from really understanding gravitational waves and even further from quantizing them; but some progress has been made, and thus we are nearer to being able to demonstrate the unity of physics.

BIBLIOGRAPHY

[1] EINSTEIN, A. 'Meaning of Relativity' (fourth edition). Princeton University Press, Princeton. 1950.

[2] *Idem.* 'Principle of Relativity' (fifteenth edition). Methuen & Co. Ltd, London. 1954.

[3] RUSSELL, Bertrand. 'A.B.C. of Relativity'. George Allen and Unwin Ltd, London. 1958.

[4] COULSON, C. A. 'Waves' (seventh edition). Oliver and Boyd Ltd, Edinburgh. 1955.

[5] LANDAU, L. D. and RUMER, G. B. 'What is Relativity?' (translated by N. Kemmer). Oliver and Boyd Ltd, Edinburgh. 1960.

The green flash and kindred phenomena

D. J. K. O'CONNELL

When the Sun is near the horizon it shows phenomena not seen at any other time. Perhaps the most dramatic of these is the transitory green flash that may sometimes be seen. Once suspected of being a physiological phenomenon, colour photography has shown the green flash to be objective: its appearance can be explained by the ordinary laws of optics. The phenomena depend not only on conditions near the ground but on those in the upper atmosphere, and may be helpful in investigating the latter.

Most voyagers in tropical seas have memories of glorious sunsets, with ever-changing colours spread over a vast expanse of sky. Some may have been fortunate enough to glimpse a fleeting effect when, at the moment of sunset, the last tip of the sun turns a brilliant green. This is the green flash, or the green ray as it is sometimes called. The first seems the more appropriate name, for what is seen is a thin horizontal green strip. A succession of green flashes may, however, appear to the naked eye as a ray shooting up from the sun (figure 2).

One need not go to the tropics to see the green flash, for it can be seen in any latitude, and over the land as well as over the sea. It is also sometimes visible at sunrise (figure 5). The flash is not necessarily green, but may even be blue or violet; Lord Kelvin [1] described a blue flash that he had seen in 1899 as the sun rose over Mont Blanc. Some years previously he had seen the green flash at sunset, and he wanted 'to see the earliest instantaneous light through very clear air, and find whether it was perceptibly blue'. He continues: 'I therefore resolved to watch for an hour till sunrise, and was amply rewarded by all the splendours I saw. . . . In an instant I saw a blue light against the sky on the southern profile of Mont Blanc, which in less than one-twentieth of a second became dazzlingly white. . . .'

The earliest known published account of the phenomenon is that given by D. Winstanley [2] in 1873. 'The green ray . . . begins at the points or cusps of the visible segment of the sun, and when the setting is nearly complete, extends from both cusps to the central space between, where it produces the momentary and intense spark of coloured light visible to the unaided eye'. This is an excellent description of the way in which the green flash has often been seen at my observatory. Winstanley quotes from a letter of Joule, describing his own observations of the flash and still earlier ones by Baxendell. The earliest dated observation that I have been able to trace was made by Swan in 1865 [3]: from the Righi at sunrise he saw a dazzling emerald-green flash over a distant mountain. He did not, however, publish this observation until 1883, perhaps after reading Jules Verne's romance *Le rayon vert*, which appeared in 1882 and which seems to have aroused widespread interest in the subject. At any rate, a great deal has been written since [4–6].

Opinion has been divided about the nature of the green flash. Many held it to be merely a subjective phenomenon, resulting from fatigue of the retina causing the complementary colour to be seen after the eye had been dazzled by the reddish rays of the setting sun. Rather curiously, this was the opinion of Swan after making the observation mentioned above, although it would seem evident that a green flash at sunrise cannot be the result of retinal fatigue; Swan blamed the reddish colour of the dawn sky.

On the other hand, many scientists believe that the green flash is an objective phenomenon [7]. I have myself been firmly of that opinion ever since I first saw the flash, at sunset over the Mediterranean, from the window of my office at Castel Gandolfo, and it seemed worth while trying to settle the controversy by photographing the flash in colour. The Vatican Observatory at Castel Gandolfo is well situated for such an experiment: there is an unobstructed view across the Roman Campagna to the Mediterranean, and the telescopes can be pointed right down to the horizon fifty miles away. It proved no easy task, but—thanks to the skill and patience of C. Treusch, S.J., the instrument-maker of the Observatory—many successful photographs have been taken not only of the green flash at sunset over sea and over land, and at sunrise, but also of various kindred phenomena [6].

THEORY OF THE GREEN FLASH

The main factors that are responsible for the green flash are the dispersion and absorption of the sun's light in the earth's atmosphere. The velocity of light decreases as the density of the air increases, so that a ray of light is refracted in its passage through the atmosphere, unless its source is exactly in the zenith. The ray is tilted towards the vertical, as compared with its original direction and consequently the sun or a star appears higher in the sky than it really is. The refraction increases with increasing distance from the zenith, and rays of shorter wavelengths (violet, blue, green) are bent more than those of longer wavelength. Thus the net result is that the image of a star is drawn out into a spectrum, with blue or green above and red below, the length of the spectrum increasing with decreasing altitude. This effect is shown very clearly in a photograph of Venus taken when it was near the horizon (figure 7): a series of images of the planet in different colours can be seen. The sun, when near the horizon, has a narrow green rim above and a red rim below, neither visible to the naked eye (figure 4).

The air not only refracts light but absorbs it. As a star sinks, its light passes through an ever-increasing thickness of air; the absorption increases and the star becomes steadily fainter. The absorption is not, however, uniform over the whole spectrum. Certain normal constituents of the atmosphere, such as water vapour, oxygen, and ozone, absorb some wavelengths more than others. The result of this is that certain colours fade more rapidly than others as a star nears the horizon. The absorption bands due to water vapour, often called Brewster's bands, after their discoverer, affect the orange and yellow rays very strongly, and sometimes, when the sun is setting, the orange and yellow light almost or quite disappears, so that there is a gap in the spectrum and an abrupt transition from red, or orange-red, to green.

Another factor that influences the flash is the scattering of sunlight by the molecules of the air. Because of this scattering, which also explains why the sky is blue, the violet and blue content of the sun's light grows fainter and fainter as the sun nears the horizon. When the rest of the sun has passed below the horizon, only a narrow green strip is left, the blue and violet being no longer visible. Under these conditions a green flash may be seen at the last moment of sunset. The flash may even appear completely detached from the sun's disc (figure 3) and float above the horizon after the sun has disappeared. At high altitudes, and when the air is very clear, the blue may still be visible at sunset and the flash may be blue-green, or pure blue, or, much more rarely, violet.

Dust or haze can produce a general absorption of sunlight, as well as selective absorption or scattering of particular wavelengths. In general, dust or haze makes the appearance of a green flash less likely, especially when the sun is much reddened. On the other hand, certain conditions may increase the intensity or duration of the flash, and thus make it easier to see. When the temperature of the air increases with height, what is called an inversion layer is formed in the atmosphere, and it is under such conditions that mirages occur. J. Evershed [8], on a ship in the tropics, frequently saw a brilliant green flash at sunset, and noted that on each occasion a mirage effect was conspicuous, caused by a thin inversion layer near the sea. 'At sunset the last segment of the disappearing limb was . . . reflected and reversed, causing a lenticular shape with the cusps raised about 1′ above the horizon. The green flash occurred when the green-edged cusps coalesced into a single bright patch. . . . It seems to me evident . . . that the mirage layer greatly intensifies the ordinary dispersion effect, by adding the light from the reflected image to the direct image at the moment of setting.' Reflections in inversion layers can be seen in figure 1. Experimental confirmation of the increase in intensity of the green flash due to mirage conditions has been provided by T. S. Jacobsen [9], who photographed the spectrum of the green flash and compared the intensity of the green light in it with that of the green in the spectrum of the low sun.

The duration of the flash depends on the rate at which the sun is rising or setting; this rate varies with the season. The vertical width of the green rim is about ten seconds of arc, which, in middle latitudes, would take only a fraction of a second to pass below the horizon. Atmospheric conditions may increase the width of the green rim, with a corresponding increase in the duration of the flash. At Castel Gandolfo the flash has often been seen to last one or two seconds or more, while our measures of the width of the green rim have reached, and at times exceeded, 30 seconds of arc.

The sun sinks more slowly, and the green flash lasts longer, the nearer one is to the poles. At Hammerfest (79°N), at one period in the year, the flash may last as long as 14 minutes, seven minutes during the slow sinking of the sun below the horizon and another seven minutes at sunrise,

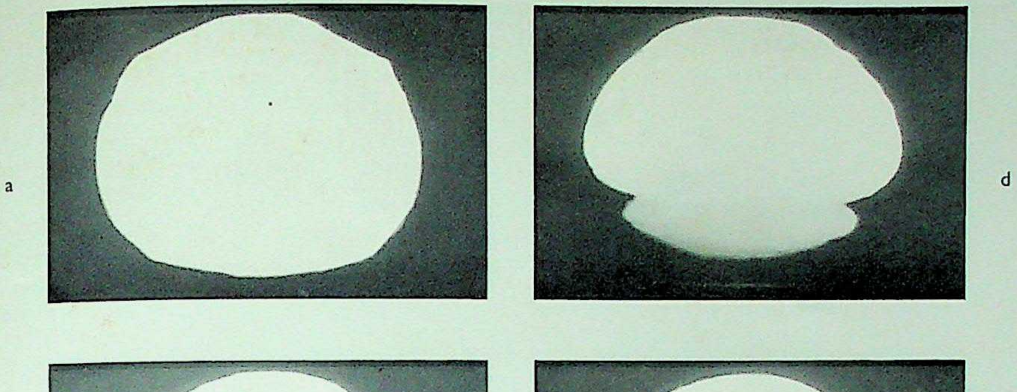


FIGURE 1 – *This picture was taken with Robot camera (aperture 7 cm, focal length 100 cm) on 15th September 1956, to show what can be achieved with a comparatively short focus. The distortions of the sun's disc are well shown, but it is not possible to obtain, with a focal length of only one metre, reproducible colour photos of the green flash. The effects of a 'blind strip' are, however, well shown. A small red flash could be seen with the naked eye in* e. *Some three minutes later a green flash appeared at the same point.*

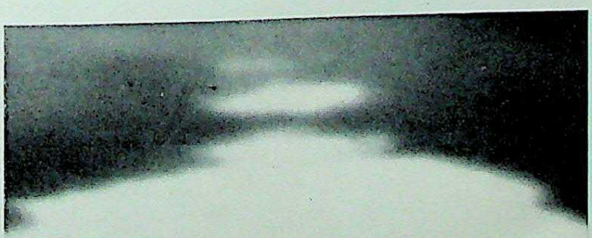


FIGURE 2 – *Multiple green flash at sunrise. A series of green flashes shooting out of the sun's rim, as seen here, can, when seen with the naked eye, give the appearance of a green 'ray'. Taken with 40 cm refractor.*

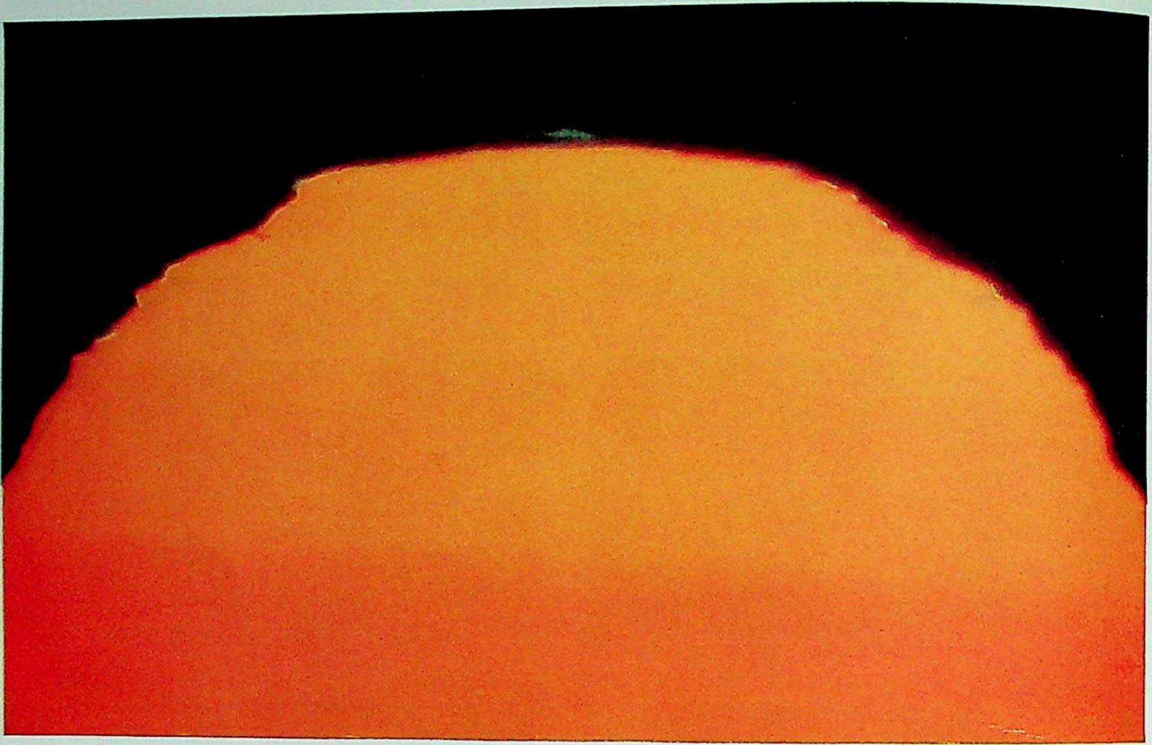


FIGURE 3 – *Green flash, shortly before sunset over the Mediterranean. (Taken on Kodachrome with 60 cm reflector.)*

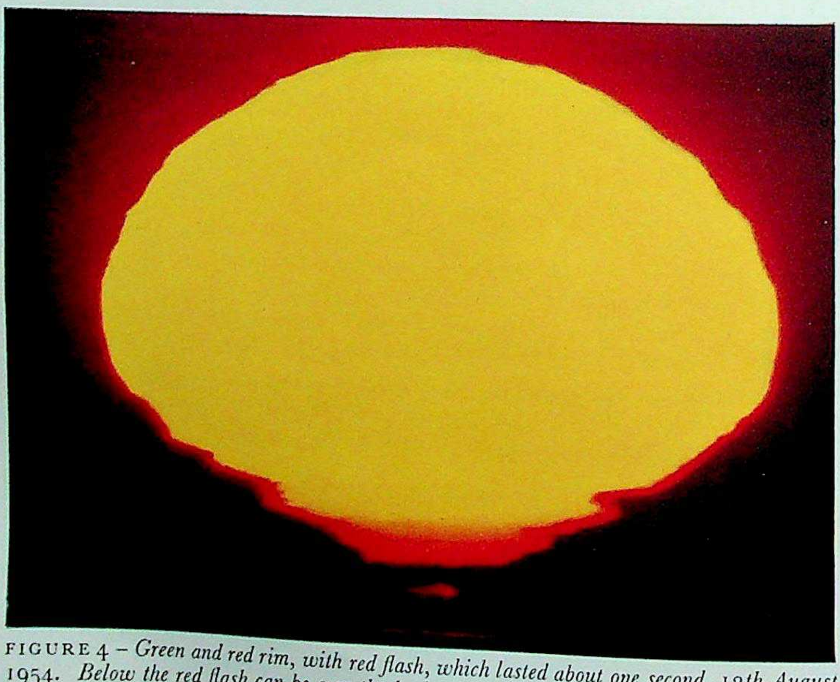


FIGURE 4 – *Green and red rim, with red flash, which lasted about one second, 13th August 1954. Below the red flash can be seen the horizon, a dark red line caused by the reflection of the red rim in the sea. (Taken on Ferraniacolour Reversal with 60 cm reflector, stopped down to 10 cm.)*

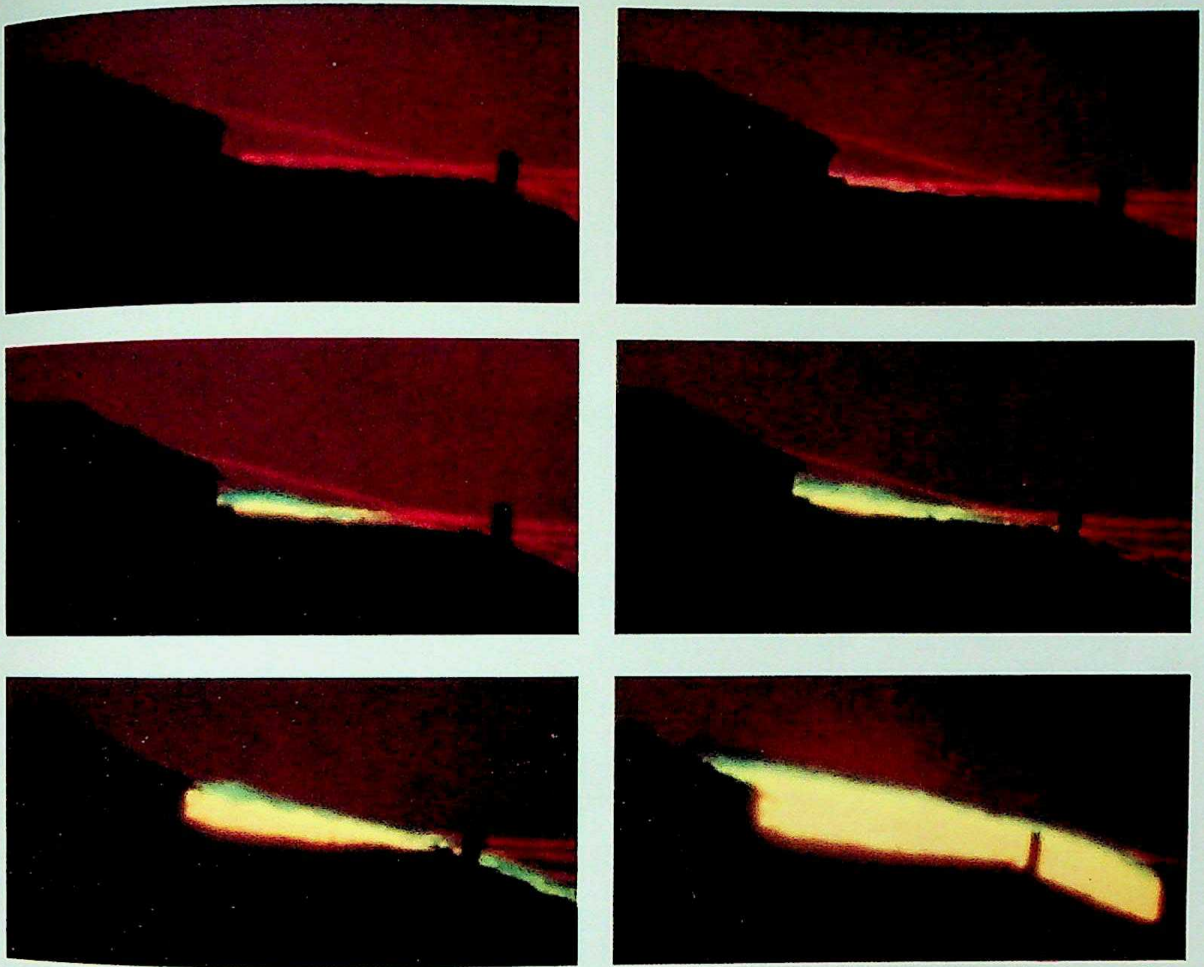


FIGURE 5 – *Green flash seen at sunrise behind medieval castle at Rocca Priora, 16th July 1956. (Taken on Ektachrome with Robot camera, 3 pictures per second, 1/250th sec. exposure, with 40 cm lens stopped down to 10 cm.)*

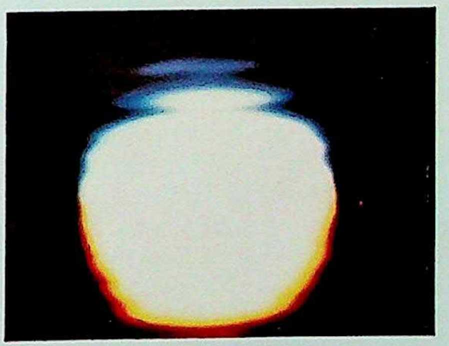


FIGURE 6 – *Artificial blue and red flashes. (Taken with Zeiss-Tessar on Ektachrome B, as described in text.)*

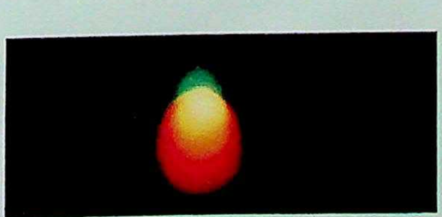


FIGURE 7 – *Venus setting. (Taken on Ektachrome Daylight with 40 cm refractor, exposure 10 sec. Magnification ×23·7.)*

following immediately on sunset [10]. A further increase is obtained when the sun at rising or setting follows the slope of a mountain. M. Minnaert [11] was able to keep the green flash in view by moving along a slope at the appropriate speed. The longest duration on record is that reported by Byrd's Antarctic expedition at Little America (78°S) in 1929, when the sun grazed the irregular horizon of the barrier ice and the green flash was seen on and off for 35 minutes. There was a strong temperature inversion near the ice, which no doubt increased the intensity.

As explained above, refraction causes the sun to appear higher in the sky than it really is. At sunset, the sun's image is raised so much that it is seen for an appreciable time when it is really below the horizon; in other words, sunset is delayed by refraction, just as sunrise is earlier for the same reason. At sunset over the Mediterranean, as seen from Castel Gandolfo, the sun is raised by refraction on the average about 1·3°, about 2·5 times the sun's diameter, and sunset is delayed by several minutes. The degree of refraction varies, depending on atmospheric conditions; it may happen that the refraction is appreciably greater than normal. In that case, sunset, besides being delayed more, lasts longer than usual, and the duration of the flash is also increased. Remarkable instances of abnormal refraction are on record. The most striking case is that reported by Dutch mariners at Novaya Zemlya in 1597. They saw the sun for fourteen days at a time when it should have been below the horizon, the refraction being more than 4°. This report was studied by Kepler and others, and many came to the conclusion that the observation was impossible and that there must have been some confusion in the dates, owing to the recent Gregorian reform of the calendar. The observations were, however, quite correct, as has been shown recently by S. W. Visser [12], who also points out that Sir Ernest Shackleton noted a similar phenomenon in the Weddell Sea during his last Antarctic expedition in 1915.

Since the formation of the green flash may be influenced by so many factors, it is not surprising that its appearance seems quite capricious. R. G. Aitken, for example, saw it at sunset from the Lick Observatory quite regularly for a long period, but at other times looked for it in vain [13]. It has been observed from many parts of the earth, but one is more likely to see it in tropical seas, or on high mountains, or in the clear air of Egypt. There is some reason to think that it was familiar to the ancient Egyptians [14]. It can indeed be seen very frequently in Egypt at sunset, and even more easily at sunrise, for the air is then particularly clear and free from dust.

The green flash is so extremely narrow that it is below the resolving power of the human eye; it can be seen with the naked eye only when the sun's disc has disappeared and the flash remains for an instant above the horizon. It is much easier to see if one uses binoculars or a small telescope, though one must, of course, be extremely careful not to injure the eyes when so doing. A neutral filter between the eyepiece and the eye can be used to reduce the intensity of the light without altering its colour; the purpose of using an optical aid is to increase the magnification. One can often see the green flash quite well with small binoculars, even when it is not visible to the naked eye. If, however, one wishes to distinguish the many fine details, such as have often been photographed here, it is necessary to have a telescope of fairly long focal length.

THE VATICAN OBSERVATORY PHOTOGRAPHS

Many have tried to photograph the green flash in colour with a short-focus camera. When the attempt failed, they concluded that the phenomenon was subjective. The Vatican Observatory photographs were taken with the Zeiss 60 cm reflector (focal length 240 cm) or with the Zeiss 40 cm visual refractor (focal length 600 cm). At first we used the reflector, in order to make sure that no colour effects were introduced by the optics. Later experiments showed that pictures taken with the refractor were free from spurious colour effects, and this telescope was used for most of the later photographs.

The first colour photographs were taken on Kodachrome film; it was a tedious business, for at that time the film had to be sent overseas for processing. As soon as Ektachrome film became available it was used almost exclusively, for it has the very great advantage that it can be developed on the spot. Much experiment was necessary to find the correct exposure, for the intensity of the light is changing and the effects to be photographed are so very fugitive. Correct exposure-time means much more than just recording the phenomena, for every effort was made to ensure that the colours in the film corresponded closely to those seen visually at the same time in the telescope.

During most of the year the sun, viewed from Castel Gandolfo, sets over the Mediterranean, and most of the sunset pictures of the green flash were taken then. Successful pictures of the flash have also been taken at sunset over land, when, during part of the summer, the sun sets behind the Tolfa mountains, sixty miles away [15]. Recording the green flash at sunrise (figure 5) presents a much more difficult task. The telescope must be pointed precisely where the flash may appear, and one must estimate beforehand the exact instant of appearance and the correct exposure. To make matters more difficult still, our eastern horizon, the Alban Hills, is rather high and close, and most of the horizon line is covered by forest. Nevertheless, pictures of the flash have been obtained. Interesting photographs have also been obtained of the blue and green rim of the sun rising behind the forests on the ridge of the hills.

There are many interesting phenomena to be observed and photographed when the sun is near the horizon. Its shape is distorted, the lower rim being raised by refraction more than the upper, so that the disc appears compressed vertically. Reflection in the sea, or in an inversion layer, gives it the form of a capital omega (Ω). There are often layers of discontinuity in the atmosphere, where the rate of decrease of the density of the air with increasing height changes abruptly. These layers cause notches to appear on the sun's rim (figure 4). An inversion layer, that is a layer of air warmer than the surface of the earth beneath it, sometimes causes a 'blind strip' which cuts the sun's disc in two (figure 1; see also [11]). At the lower edge of the blind strip there sometimes appears a red flash, and soon afterwards when the upper rim of the sun is about the same height above the horizon, a green flash may appear. Such red and green flashes have frequently been seen and photographed here (figures 1 and 4). Sometimes green strips form at the upper edge of the sun and gradually become detached, to form a green flash even before the sun sets (figure 3). At times a succession of such detached wisps or flashes have been seen here and photographs have been taken. Figure 2, taken at sunrise, shows three such green flashes, one above the other. Much more rarely seen than the green flash is a red flash below the sun. The red flash has been photographed here several times (figure 4). Venus presents an interesting sight when setting, as can be seen from figure 7.

It is interesting to study the effects of scintillation on the sun's rim: the changes are so rapid that it requires a cine camera to follow them. Successful photographs of scintillation effects, and of the development of detached green strips, have been made with a cine camera (without lens, of course), attached to the 40 cm refractor. The long focal length of this telescope makes possible the recording of very fine details that would be lost with a shorter instrument.

Figure 6 shows the results of an experiment to produce artificial blue and red flashes. In this, a glass mirror was placed on the ground under a flint-glass objective prism of 60 cm diameter and 8° refracting angle. The light from a ceiling lamp, an opal sphere of 20 cm diameter, passed through the prism, was reflected by the mirror and passed back through the prism to a camera; this simulated very well the image of the sun when near the horizon. To reproduce the effect of discontinuity layers, three panes of window glass were set above the prism, slightly inclined to the horizontal.

In conclusion, it is worth emphasizing that the appearance of the green flash, and of other phenomena referred to here, is dependent on the state of our atmosphere, not only on the dense lower layers of the air but on conditions at great altitudes. Just now, when men are undertaking their first extra-terrestrial voyages, it is essential to know all that we can about the outer atmosphere. We are indeed learning much by means of rockets and artificial satellites, but there is still a great deal that we do not know about these very tenuous upper regions of the air. It may well be that studies of such phenomena as the green flash may contribute useful information.

REFERENCES

[1] KELVIN, Lord. *Nature, Lond.*, **60**, 411, 1899.

[2] WINSTANLEY, D. *Proc. Manchr. Lit. Phil. Soc.*, **13**, 1, 1873.

[3] SWAN, W. *Nature, Lond.*, **29**, 76, 1883.

[4] FISHER, W. J. *Pop. Astr.*, **29**, 251, 382, 1922.

[5] KUIPER, P. F. 'De Groene Straal', Helder. 1926.

[6] O'CONNELL, D. J. K. 'The Green Flash and Other Low Sun Phenomena'. Vatican Observatory. North Holland Publishing Co., Amsterdam; Interscience Publishers Inc., New York. 1958.

[7] See, for example, DANJON, A. and ROUGIER, G. *Ann. Obs. Strasbourg.*, **1**, 105, 1926.

[8] EVERSHED, J. *Nature, Lond.*, **111**, 13, 1923.

[9] JACOBSEN, T. S. *J. R. astr. Soc. Can.*, **46**, 93, 1952.

[10] GUGLIELMO, G. *R. C. Accad. Lincei*, (5), **25**, 417, 1916.

[11] MINNAERT, M. 'Light and Colour in the Open Air' (2nd ed.), p. 59, Dover Publications, New York. 1954.

[12] VISSER, S. W. *Proc. R. Acad. Sci. Amst., B*, **59**, 375, 1956.

[13] AITKEN, R. G. *Leafl. astr. Soc. Pacif.*, **3**, 177, 1939.

[14] See [6], p. 8.

[15] See [6], plate XI.

Optical rotatory dispersion

CARL DJERASSI

The variation of optical rotation with wavelength has been known for nearly one hundred and fifty years, but its application to organic chemical and biochemical problems started only recently. Measurements of anomalous optical rotatory dispersion have given a powerful new method for the solution of various structural, stereochemical, and kinetic problems. In particular, rotatory-dispersion studies were found to be almost uniquely useful for determining the absolute configuration of organic compounds and in the detection of subtle conformational changes. Some of these recent developments are here discussed.

The organic chemist, especially the investigator of natural products, has always been very ready to adopt new physical methods for the solution of his problems. Today, it is difficult to imagine modern organic chemical research being conducted without the use of ultraviolet, infrared, or nuclear magnetic resonance, spectroscopy, and it is profitable to consider for a moment how these tools became acceptable to the organic chemist. In every case the basic principle was discovered by a physicist or physical chemist, who was usually concerned only with the phenomenon itself, rather than with its application. It was used in the organic chemical laboratory only when instrumentation became sufficiently advanced, or simplified, for a relatively large number of measurements to be made with organic compounds. These measurements were then empirically correlated with some structural feature of the organic molecule, and this invariably led to two developments. One, a rapid acceptance of the method by organic chemists in general, with a consequent enormous increase in the number of measurements, and hence refinements in correlation. The other, a revival or extension of interest in the theoretical aspects, and a more fundamental explanation of the semi-empirical discoveries of the practical chemist.

This has been the history of virtually all physical methods that have found a temporary or a permanent place in organic and biochemical research; the only difference between them has been the time-lag between the initial physical discovery and the first widespread use by the organic chemist. In the field of infrared spectroscopy, for instance, the interval amounted to approximately forty years, but less than a dozen years passed between the discovery of nuclear magnetic resonance and its extensive use in organic chemistry.

HISTORY

The development of optical rotatory dispersion followed this general pattern, except that nearly one hundred and forty years elapsed between the original discovery by J. B. Biot [1] of the change of optical rotation produced by quartz with wavelength and the application of this phenomenon to organic chemical problems [2]. In the interval, considerable research was conducted, largely by physical chemists, and reviews by three of the pioneers, T. H. Lowry, P. A. Levene, and W. Kuhn [3], show on the whole a rather understandable preoccupation with theoretical aspects. During this entire period, up to about 1954, considerably less than one hundred optical rotatory-dispersion curves had been recorded in the ultraviolet region of the spectrum, principally because of the instrumental difficulties. Indeed, between the 1930s, when experimental work by the research groups led by these three men had terminated, and 1955, only a few publications appeared on optical rotatory-dispersion measurements in the ultraviolet, and there are practically no records of the use of this method to solve an organic or biochemical problem. This contrasts with two striking indications of present activity and interest in optical rotatory-dispersion, with special application to organic chemistry and biochemistry. The advances in instrumentation are illustrated by the fact that, in the author's laboratory, approximately three thousand ultraviolet rotatory-dispersion curves were measured in less than eight years. A quantitative measure of the general acceptance of optical rotatory-dispersion as a physical tool is afforded by the fact that, during the past eighteen months, approximately two hundred publications have appeared in which ultraviolet rotatory-dispersion measurements have been employed for the solution of organic chemical or biochemical problems.

GENERAL PRINCIPLES

To appreciate the reasons for this renaissance, a brief discussion of optical rotatory dispersion is required. We may consider a beam of plane-polarized light to be made up of a right and a left circularly polarized component. Although no rotation will be observed if these two components pass through a medium with equal velocities, the plane of polarization of the emergent light will be rotated if the refractive indices of the medium for right and left polarized light are unequal. This dispersive feature is the principal reason for monochromatic polarimetry as, for instance, in the conventional measurement of optical rotation at one wavelength, usually the sodium D line. In addition to unequal velocity of transmission, there may also occur unequal absorption of left and right circularly polarized light. The emergent light will then be elliptically polarized; this absorptive effect is known as circular dichroism. The combination of circular dichroism and optical rotation is named the Cotton effect, after its discoverer. It was discovered in 1895. As will be shown, it is the experimental measurement of the Cotton effect, and its correlation with certain structural and stereochemical features, that is chiefly concerned in the application of optical rotatory dispersion in organic chemistry.

In principle, the experimental determination of an optical rotatory-dispersion curve is a simple matter that involves the measurement of the optical rotation of a given substance at different wavelengths, rather than just at the conventional sodium D line. The apparatus consists essentially of a light source; a monochromator producing light of a given wavelength; a polarizer; a cell containing the optically active material in solution; an analyser; and a device for measuring the angle of rotation. The polariser and analyser are usually quartz Rochon prisms, and the angle of rotation is measured visually when working with visible light. However, most organic compounds are colourless, and absorb only in the ultraviolet; consequently, if one wishes to measure the Cotton effect of colourless substances, the determinations must be carried out with ultraviolet light. It is here that experimental difficulties arise. All the earlier work [3] in the ultraviolet region used laborious photographic procedures, and it was only the commercial development of a spectropolarimeter [2] that made possible the relatively rapid and simple measurement of a rotatory-dispersion curve in the ultraviolet. The measurement could be made in an hour or two by this method, whereas the photographic one needed several days.

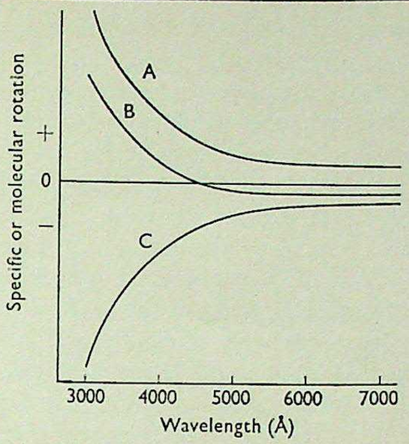


FIGURE 1 – *Typical plain rotatory-dispersion curves (A and B, positive; C, negative).*

There are two types of dispersion curves [4]. Figure 1 illustrates some plain curves, that is, dispersion curves that exhibit no 'maxima' or 'minima'; in these the rotation simply increases as one proceeds into the ultraviolet. Curves A and B of figure 1 are called positive plain curves, even though curve B starts out on the negative rotation side in the visible and crosses the zero rotation axis in the ultraviolet; curve C is called a negative plain curve.

An anomalous optical rotatory-dispersion curve is one that exhibits rotational maxima and minima; an idealized example is reproduced in figure 2. Curve A represents a single positive Cotton effect; Curve B is typical of a negative Cotton effect, since the first extremum in going from the visible to the ultraviolet is a rotational minimum. The rotational maxima and minima are called peaks and troughs, so as not to confuse them with ultraviolet absorption maxima and minima. Thus the positive Cotton effect of curve A in figure 2 would be exhibited by an optically active substance with a chromophore absorbing at wavelength λ_0.

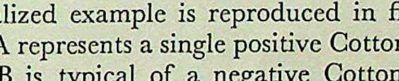


FIGURE 2 – *Typical anomalous rotatory-dispersion curves (A, positive Cotton effect; B, negative Cotton effect).*

This represents approximately the midpoint, in terms of wavelength units, between the positions of the peak and the trough of the rotatory-dispersion curve. This intimate relationship between the absorption spectrum and the optical rotatory-dispersion of a given substance will usually indicate whether it will exhibit an anomalous rotatory-dispersion curve in a given wavelength region.

Plain curves of the type shown in figure 1 will be exhibited generally by optically active substances, such as hydrocarbons, alcohols, and carboxylic acids, that do not absorb in an accessible region of the spectrum; an anomalous curve will be shown by a substance containing a chromophore, such as a carbonyl or nitro group, that does absorb in such a region. In reality, every plain curve will eventually become anomalous as it approaches its optically active absorption band; but in many compounds these bands may occur so far down in the ultraviolet, or the compounds may absorb with such intensity, that rotation measurements are not feasible in that spectral region. To illustrate this point, the anomalous rotatory-dispersion curves shown in figure 2 are actually plain curves over the range 7000–4500 Å and become anomalous only as they approach the region of the absorption maximum ($\lambda_0 = 3000$ Å).

For substances whose behaviour is represented by either figure 1 or figure 2, the optical rotations will be much greater in the ultraviolet than in the visible. The historical choice of the sodium D line for monochromatic rotation measurements is therefore a rather poor one. Much smaller amounts of optically active substrate are required for rotation measurements in the ultraviolet than in the visible, and most of the ultraviolet rotatory-dispersion measurements in our laboratory have been conducted with 0·1–1·0 mg of material, a quantity that is usually much too small for rotation measurements at the sodium D line.

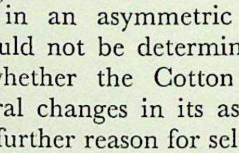
Between 1812 and 1954 only approximately fifty Cotton-effect curves had been measured experimentally [3], and the various chromophores that had been investigated included a few ketones, such as camphor, whose rotatory-dispersion curve was similar to curve A in figure 2. While the availability of this experimental material did not lead to any widespread applications, it nevertheless demonstrated the feasibility of measuring Cotton effects, provided that substances with suitable chromophores were selected. When the first model of a commercially available spectropolarimeter, the Rudolph spectropolarimeter, was delivered to our laboratory, we therefore set out to see if the Cotton effect could be related to structure.

APPLICATIONS TO STEREOCHEMISTRY

Although for stereochemical correlations [5] plain rotatory-dispersion curves are preferable to single rotation measurements at the sodium D line, especially when the curves cross the zero line, they are obviously limited in scope. On the other hand, if variations in the asymmetric environment of a given spectroscopically accessible chromophore should result in Cotton-effect curves of different shape or sign, and if these differences could be related to certain structural or stereo-chemical features of the molecule, a valuable physical tool would be available to the organic chemist.

We selected the carbonyl chromophore for initial experimental examination because it absorbs in an accessible spectral region (2800–3300 Å) with a rather low extinction that would permit rotation measurements through this region of maximal absorption; this is essential if Cotton effects are to be determined experimentally. Furthermore, earlier work [3] on camphor and a few related terpene ketones had shown that such ketones exhibit Cotton-effect curves—in other words, that the intrinsically symmetrical carbonyl chromophore $>C{=}O$ became optically active when placed in an asymmetric environment, although it could not be determined from these few studies whether the Cotton effect would reflect structural changes in its asymmetric surroundings. A further reason for selecting the carbonyl chromophore was the ready availability of many optically active ketones and aldehydes, or at least alcohols from which they could be obtained by oxidation.

With the choice of the chromophore—that is, the absorptive component of the rotatory dispersion—settled, it was necessary to decide upon a reasonable asymmetric framework that would offer sufficient scope for selective alteration to see whether such changes would result in variations in the shape or sign of the Cotton-effect curve. For this purpose we chose a uniquely suitable group of compounds, the steroids, which are characterized by the following skeleton:



The advantages of this group are manifold. All steroids are optically active, of known absolute configuration, and usually of fixed and known conformation. Furthermore, probably several thousand representatives of this class have been prepared during the past thirty years. Finally, most of the steroids possess a chain of seven contiguous asymmetric carbon atoms (circled), so that irrespective of where a carbonyl group is placed in this molecule, it can, at most, be only one carbon atom away from this asymmetric chain. By placing a carbonyl group in the eleven possible locations of the four-ring skeleton, the chromophore will find itself in different asymmetric situations. The incorporation of a trigonal carbon atom into any one of the rings will also cause a displacement of the valency angles, with a concomitant ring-geometry distortion that may make itself felt through varying ranges of this asymmetric backbone. As we had in our laboratory many such steroids, and also a modern spectropolarimeter, we could measure, in the ultraviolet, many hundred rotatory-dispersion curves of closely related ketones. This initial work [6] demonstrated that variations in the shape and sign of the Cotton effects can indeed be related to many important molecular parameters. This prompted us to carry out a much more extensive investigation of the application of rotatory-dispersion measurements in organic chemistry than had been contemplated initially. Much of this work has appeared between 1955 and 1960 in some forty articles from our laboratory, and it has been summarized recently in book form by the present author [2]. For this reason, only a few highlights will be mentioned below, to indicate the general scope of the method and future directions of research.

Figure 3 shows a typical example of the structural information that can be derived from a rotatory-dispersion curve. It contains the dispersion curves of three isomeric cholestanones, distinguished only by the location of the carbonyl group; it will be noted that they differ significantly either in shape or in sign. A point of great importance is that these characteristic curves persist in the presence of many other substituents, such as esters, acids, alcohols, and lactones, that do not absorb in the same spectral range as the carbonyl group. Consequently we have now a tool that can frequently tell us, using less than 1 mg of recoverable substance, where a keto group, and hence the corresponding hydroxyl function, is located in an unknown steroid. At present, no other physical method will give this information.

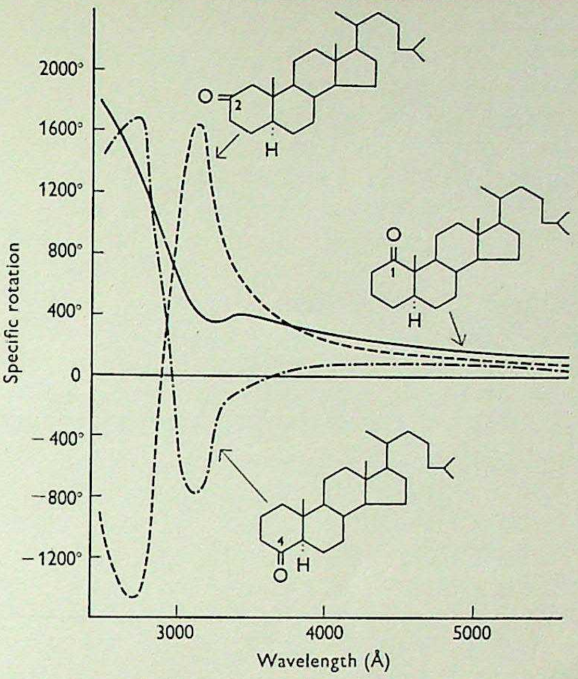


FIGURE 3 – *Optical rotatory-dispersion curves of some isomeric cholestanones.*

The differences in the rotatory-dispersion curves of figure 3 were brought about by moving the carbonyl group around the asymmetric framework and thus placing it each time in a distinct asymmetric environment; this change is reflected in the dispersion curve. Alternatively, such a change in environment can be produced by keeping the chromophore in the same place and altering the stereochemical environment. Thus, reduction of the double bond of the male sex hormone testosterone yields two isomers, 5α- and 5β-androstan-17β-ol-3-one, differing only in the nature of the A/B ring juncture. As shown in figure 4, the 5α-isomer exhibits a positive (A) and the 5β-isomer, a negative (B) Cotton effect. These characteristic rotatory-dispersion features remain unchanged, even if the substituents at C-5 or C-10, or both, in these molecules are replaced by other 'non-chromophoric' groups, such as CH_2OH, CN, or CO_2CH_3; this has proved to be very useful in the steroid field for determining the nature of the A/B ring juncture in an unknown 3-keto steroid.

When the carbonyl group is placed next to the ring juncture, equilibration between the two isomers is possible in the presence of acid or base. This is illustrated in figure 5 for the two 15-ketones of the bile-acid series, which differ only in the configuration at C-14. The C/D *cis* isomer (B), with a negative Cotton effect, is the more stable isomer, and the kinetics of the isomerization can

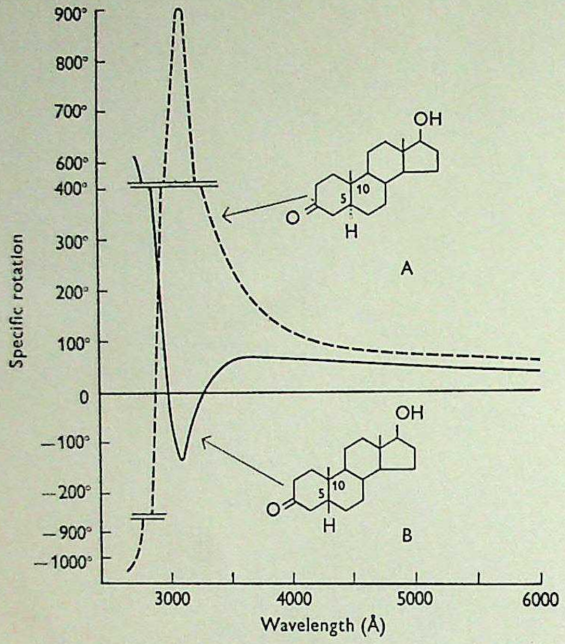


FIGURE 4 – *Optical rotatory-dispersion curves of 5α-androstan-17β-ol-3-one (dihydrotestosterone) (A) and 5β-androstan-17β-ol-3-one (B).*

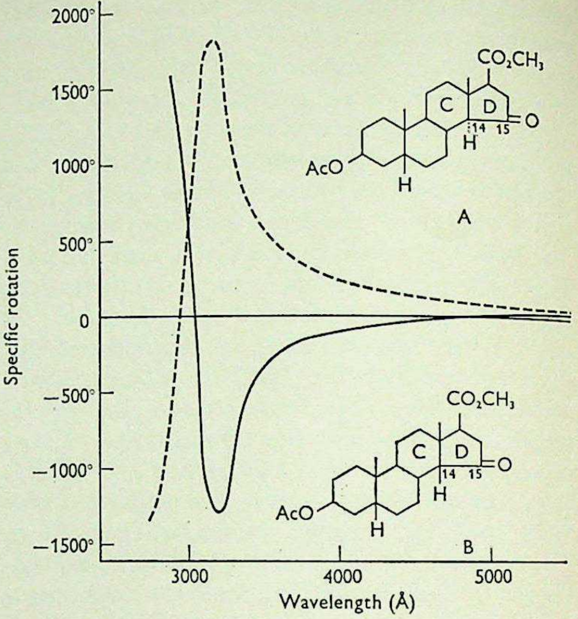


FIGURE 5 – *Optical rotatory-dispersion curves of methyl 3β-acetoxy-15-oxoetianate (A) and methyl 3β-acetoxy-15-oxo-14β-etianate (B).*

be determined very conveniently on a micro-scale by taking advantage [7] of the information contained in figure 5. It will be noted that the rotations of the two isomers are rather similar at the sodium D line in the visible, whereas in the ultraviolet near 3200 Å the difference in specific rotation amounts to over 3000°. All that is necessary, therefore, is to measure the rate of change in rotation at some convenient wavelength in the ultraviolet, and to calculate from these data the kinetics and the position of equilibrium.

Figures 3, 4, and 5 represent typical examples from the steroid series of structural, stereochemical, and analytical-kinetic applications of dispersion measurements. That these generalizations from the steroid group are, in fact, of much wider scope was demonstrated [8] by a process of dissection, by which it was found that only the immediate asymmetric bicyclic environment of the carbonyl group is responsible for the characteristic rotatory-dispersion features. This conclusion led immediately to the two most important applications of rotatory dispersion, namely the assignment of absolute configurations and the detection of conformational alterations.

APPLICATIONS IN CONFORMATIONAL ANALYSIS

The use of this tool in conformational analysis is illustrated in figure 6. *Trans*-10-methyl-2-decalone (I) represents the bicyclic analogue of a 5α-3-keto steroid (A in figure 4), and its conformation can be defined unambiguously as the all-chair form IA (figure 6). The rotatory-dispersion curve of this ketone I proved to be practically identical with that of steroid A in figure 4. A considerably more difficult problem concerns the isomeric *cis*-10-methyl-2-decalone (II in figure 6), where two all-chair conformations (IIA and IIB) have to be considered. Of these, IIA is called the steroid conformation, as steroids are frozen in this conformation by virtue of the additional ring fusions. We had already established [8] that the more distant rings in the steroid series do not play an important effect as far as the typical shape of the rotatory-dispersion curve of a given ketone is concerned; the negative Cotton effect of the steroid B in figure 4 will thus serve as a suitable reference curve for the steroid conformation of the bicyclic ketone II in conformation IIA (figure 6). It will be seen that, in the non-steroid conformation IIB, the carbonyl group finds itself in a different asymmetric environment, and its rotatory-dispersion curve would therefore be expected to be quite different from that of the steroid model. When II was synthesized [8] in optically active form and its rotatory-dispersion curve was measured, it proved to be extremely similar to that of the

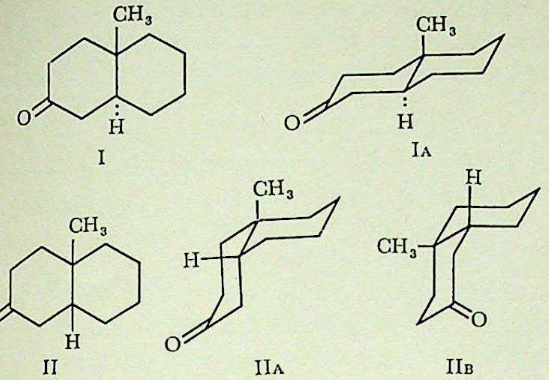


FIGURE 6 – *Conformational representations of* trans-10-*methyl-2-decalone (I) and* cis-10-*methyl-2-decalone (II).*

steroid B in figure 4. One can deduce, therefore, that the ketone II exists in the steroid conformation IIA, a conclusion which, *a priori*, was not self-evident to the stereochemist. At the present time, rotatory dispersion has been the only physical method that has shed any light on the conformation of such a decalone in solution.

THE DETERMINATION OF ABSOLUTE CONFIGURATION

The determination of absolute configuration, that is, the differentiation between mirror-image representations, has been of concern since the earliest days of stereochemistry. The only direct procedure is that of X-ray analysis [9], but so far this has been used in only very few instances. What is really needed is a quick method that permits the establishment of the absolute configuration of a given molecule by relating it to another reference compound whose absolute configuration has been settled directly or indirectly (after chemical transformation) by the X-ray method. Optical rotatory-dispersion measurements have satisfied this very important need.

In principle, the approach is quite simple [8] and is based on the same premise as was employed for examining the conformation of *cis*-10-methyl-2-decalone. We have shown that the characteristic shape of a rotatory-dispersion curve is governed by the bicyclic stereochemical and conformational environment of the chromophore—in the present instance, the carbonyl group. As additional non-chromophoric substituents do not interfere, all that is necessary is to compare the rotatory-dispersion curve of the unknown compound with that of a known reference substance possessing the same chromophore in a similar environment.

The absolute configuration of the steroids is known [10] and their rotatory-dispersion curves, therefore, can be used for such reference purposes. A typical example is reproduced in figure 7, which shows rotatory-dispersion curves of the steroid Δ^4-cholesten-3-one (I) and the sesquiterpene carissone (II). The hydroxy-isopropyl substituent at C-7 of carissone represents a non-chromophoric substituent, as does the extra methyl group at C-4. This latter point was established [8] in a separate experiment in which it was found that the rotatory-dispersion curve of Δ^4-cholesten-3-one (I) was not altered in any of its essential details by the addition of a methyl group at C-4. Consequently, the carbonyl chromophore in the steroid I of known absolute configuration and in the sesquiterpene carissone (II) must be situated in identical bicyclic environments. As their rotatory-dispersion curves are very similar in shape and sign, it is certain that the absolute configuration of carissone is that represented in formula II, rather than its mirror image. The third substance in figure 7, the tricyclic ketone III, also has the same relative stereochemical environment in rings A and B as the steroid I. Its rotatory-dispersion curve, however, is essentially of mirror-image type, from which it follows that, in terms of its absolute configuration, ketone III belongs to the antipodal series.

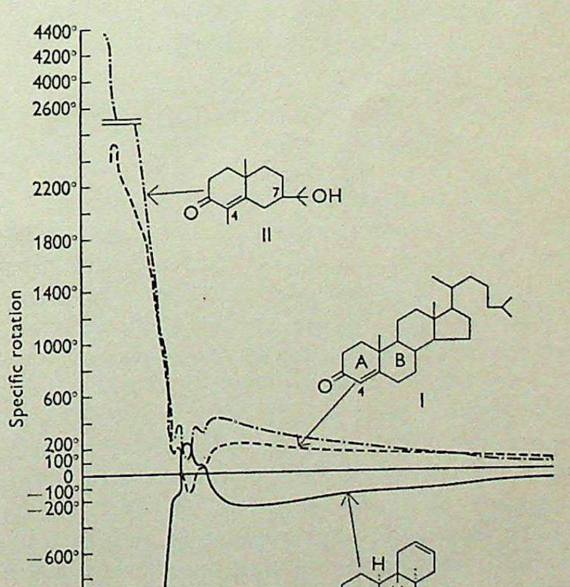
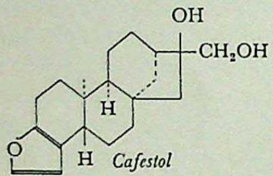
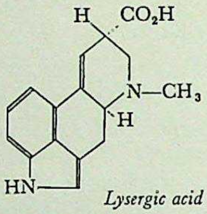
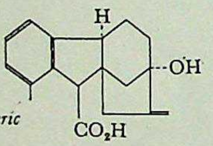


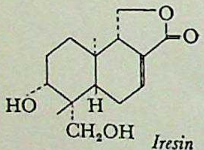
FIGURE 7 – *Optical rotatory dispersion curves of* Δ^4-*cholesten-3-one (I), carissone (II), and* (—)-1,14-*dimethyl-2-oxo-*$\Delta^{1(11),6}$-*decahydrophenanthrene (III).*

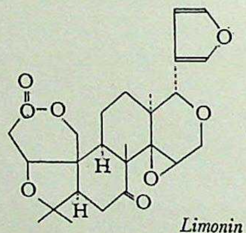
Various refinements of this approach have been introduced recently, including semi-empirical generalizations that often make it possible to predict the sign of the Cotton effect of an unknown ketone without requiring the rotatory-dispersion curve of a reference compound. These generalizations have already been reviewed elsewhere [11]. Suffice it to say that by the approach illustrated in figure 7, or by extensions of this approach [11], the absolute configurations of a wide variety of very complicated natural products have been established very quickly. A few typical examples are listed below, and it is obvious that any other classical approach would have required a tremendous effort in order to arrive at the same conclusions.


Cafestol


Lysergic acid

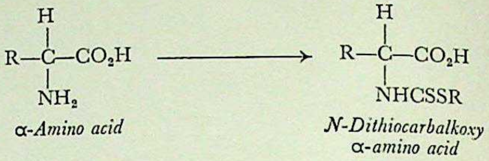

Allogibberic acid


Iresin


Limonin

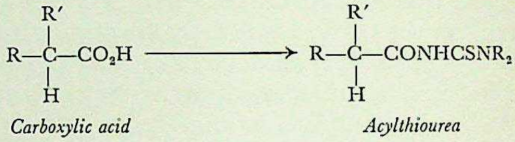
Most of the discussion until now has centred on the carbonyl chromophore. Lately, our attention, as well as that of other investigators, has been directed to other chromophores, and very substantial progress has already been achieved. Recent examples are dithiocarbamates and xanthates [12], disulphides and diselenides [13], phthalimides [14], acylthioureas [15], and osmates [16]. This partial list is important not only because it illustrates the range of different chromophores that may give rise to Cotton-effect curves, but even more significantly because it includes derivatives of important non-chromophoric functional groups.

For instance, α-amino acids do not absorb in a convenient region of the ultraviolet, and they thus show only plain rotatory-dispersion curves; however, their readily prepared N-dithiocarbalkoxy derivatives (dithiocarbamates) exhibit strong Cotton effects. In fact, the sign of the Cotton effect can be related directly [12] to the absolute configuration of the asymmetric centre, members of the L-series showing a positive Cotton effect and of the D-series a negative one.


$$\mathrm{R{-}\overset{H}{\underset{NH_2}{C}}{-}CO_2H} \longrightarrow \mathrm{R{-}\overset{H}{\underset{NHCSSR}{C}}{-}CO_2H}$$

α-Amino acid *N-Dithiocarbalkoxy α-amino acid*

Similarly, carboxylic acids do not absorb above 2100 Å, but their derived acylthioureas [15] contain a low-intensity absorption maximum near 3400 Å. It is not surprising, therefore, that such derivatives should yield Cotton-effect curves, and that the sign of these Cotton effects can be used to deduce the absolute configuration of the α-asymmetric centre of the parent carboxylic acid.


$$\mathrm{R{-}\overset{R'}{\underset{H}{C}}{-}CO_2H} \longrightarrow \mathrm{R{-}\overset{R'}{\underset{H}{C}}{-}CONHCSNR_2}$$

Carboxylic acid *Acylthiourea*

Olefins are essentially transparent above 2000 Å, but they react readily with osmium tetroxide to form osmate esters. These esters, and especially

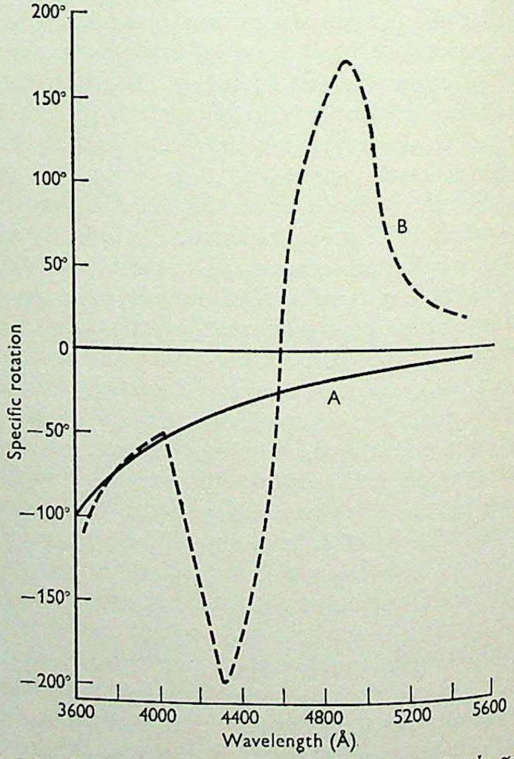


FIGURE 8 – *Optical rotatory-dispersion curve of poly-α-L-glutamic acid at pH 4·9. Curve A, polypeptide without dye; curve B, polypeptide-acriflavine complex (redrawn from Blout, E. R. and Stryer, L. [19], by permission of the authors).*

their dipyridyl complexes, have been found [16] to exhibit strong Cotton effects in a very convenient region of the spectrum. As these osmates can be prepared on a micro-scale, there is now available a means of studying certain structural and stereochemical features of olefins in terms of anomalous optical rotatory-dispersion curves.

THE CONFORMATION OF POLYPEPTIDES AND PROTEINS

One very exciting application of optical rotatory dispersion is the determination of the conformation of polypeptides and proteins. As this has been reviewed recently in great detail by E. R. Blout [17], only brief reference will be made to it here. The hydrogen-bonded α-helix structure of proteins, initially proposed by L. Pauling and R. B. Corey, is now generally accepted. One of the most effective ways of studying such a conformational feature has been by optical rotatory dispersion, even though such curves have until recently been of the plain type. Specifically, it was found that the plain curves of polypeptides in a random conformation follow a simple Drude equation (this relates optical rotatory power and wavelength outside the region of an optically active absorption-band), although this is not true of the helical conformation. The consequent mathematical generalization, largely due to W. Moffitt, has proved to be very useful. M. Goodman and collaborators [18] have demonstrated by such rotatory-dispersion measurements that the critical range where synthetic polypeptides of γ-methyl glutamic acid shift from the random to the helical structure falls between the hepta- and nona-peptide.

Two important recent observations [19, 20] indicate that anomalous optical rotatory dispersion with experimentally determined Cotton effects will play an increasingly significant role in the polypeptide and protein fields. In the first paper it was noted that while poly-α L-glutamic acid exhibits only a plain rotatory-dispersion curve (A in figure 8), its complex with the optically inactive dye acriflavine shows a Cotton effect (B in figure 8) centred around the 4570 Å absorption band of the dye. Such Cotton effects are observed with the dye-polypeptide complex only when the polypeptide is in the helical rather than the random conformation. The complexes thus offer a strikingly simple approach to the study of helical conformations, as well as to the determination of the 'handedness' of the helix. Of even more general applicability is the observation that the Cotton effects of polypeptides and proteins may be determined directly in the 2100–2300 Å region with the new automatically recording Rudolph spectropolarimeter.

REFERENCES

[1] Biot, J. B. *Mém. Inst.*, **1**, 1, 1812; *Mém. prés. Acad. Sci., Paris*, **2**, 41, 1817.

[2] Djerassi, C. 'Optical Rotatory Dispersion: Applications to Organic Chemistry'. McGraw-Hill, New York. 1960.

[3] Lowry, T. H. 'Optical Rotatory Power'. Longmans, Green & Co., London. 1935.
Levene, P. A. and Rothen, A., in H. Gilman, 'Organic Chemistry', Vol. II, Chapter 21. John Wiley & Sons, New York. 1938.
Kuhn, W. *Ann. Rev. Phys. Chem.*, **9**, 417, 1958.

[4] See chapter 2 in [2] and Djerassi, C. and Klyne, W. *Proc. chem. Soc., Lond.*, 55, 1957.

[5] See chapter 16 in [2].
Jones, P. M. and Klyne, W. *J. chem. Soc.*, 871, 1960.
Sjöberg, B. *Ark. Kemi*, **15**, 451, 1960.

[6] Djerassi, C. *et al. J. Amer. chem. Soc.*, **77**, 4354, 4359, 4364, 1955; *ibid.*, **78**, 440, 3163, 3761, 1956.

[7] Allinger, N. L., Hermann, R. B. and Djerassi, C. *J. org. Chem.*, **25**, 922, 1960.

[8] Djerassi, C., Riniker, R., and Riniker, B. *J. Amer. chem. Soc.*, **78**, 6362, 1956.
Djerassi, C. and Marshall, D. *Ibid.*, **80**, 3986, 1958.

[9] Bijvoet, J. M. *Endeavour*, **14**, 71, 1955.

[10] Fieser, L. F. and Fieser, M. 'Steroids'. Reinhold Publishing Corporation, New York. 1959.

[11] See chapters 9, 10, and 13 in [2].

[12] Sjöberg, B., Fredga, A., and Djerassi, C. *J. Amer. chem. Soc.*, **81**, 5002, 1959.

[13] Djerassi, C., Fredga, A., and Sjöberg, B. *Acta chem. Scand.* (In press.)

[14] Brewster, J. H. and Osman, S. F. *J. Amer. chem. Soc.*, **82**, 5754, 1960.

[15] Djerassi, C. and Undheim, K. *Ibid.*, **82**, 5755, 1960.

[16] Bunnenberg, E. and Djerassi, C. *J. Amer. chem. Soc.*, **82**, 5953, 1960.

[17] Blout, E. R. Chapter 17 in [2].

[18] Goodman, M., Schmitt, E. E., and Yphantis, D. *J. Amer. chem. Soc.*, **82**, 3483, 1960.

[19] Blout, E. R. and Stryer, L. *Proc. Nat. Acad. Sci., Wash.*, **45**, 1591, 1959.

[20] Simmons, N. S. and Blout, E. R. *Biophys. J.*, **1**, 55, 1960.
Simmons, N. S., Cohen, C., Szent-Györgyi, A. G., Wetlaufer, D. B., and Blout, E. R. (Unpublished observations.)

The nucleolus of the cell nucleus

J. L. SIRLIN

The nucleolus, an organelle of the cell nucleus, was first described by Fontana nearly two centuries ago, but its nature remained obscure until recently. Its precise function in the cell has still not been completely elucidated, but it is known to be involved in protein synthesis. This article discusses current theories of the biochemical function of the nucleolus.

'*On observe un corps oviforme, ayant une tache dans son milieu.*' Fontana used these words in 1781 [1] in describing cells in the slime of an eel; it is the first mention of the nucleolus; the nucleus itself, Fontana's oviform body, had been first described by Leeuwenhoek in 1702. The nucleolus is present in most cell nuclei [2]. In general it is a roundish body, numbering one or a few per nucleus, and denser than the remainder of the chromatin, in which it is embedded, because of its high protein content. The subtlety of the function of the nucleolus is such that it was not described in general chemical terms until the 1940s. This was largely a result of the development, by the schools of T. O. Caspersson in Sweden and J. Brachet in Belgium, of knowledge of the cell chemistry of ribonucleic acid (RNA) and protein. The precise chemistry of the function is still not clear.

Apart from RNA, which is its most characteristic constituent, the nucleolus contains lipids, carbohydrates, and mineral substances. Strictly speaking, it lacks deoxyribonucleic acid (DNA), the most characteristic constituent of chromosomes. However, it is surrounded, in most cytological preparations, by condensed chromatin (the distended chromosomes) forming the nucleolus-associated chromatin. Because this material is considered by some authors to be part of the nucleolus, the latter is often said to contain DNA. Sometimes the nucleolus-associated chromatin is of the type known as heterochromatin.

Inside the nucleus, the nucleolus is subject to movements. These may be either passive, due to the gel-like consistency of chromatin, or active, such as emission of prolongations, amoeboid movements, fusion with other nucleoli and the nuclear membrane, and even fragmentation; the active are the more important. This plasticity of the nucleolus may lead to very bizarre shapes, as in certain nurse cells and in ciliate protozoa, in which it departs considerably from the usual roundish appearance. The movements of the nucleolus are usually more pronounced during periods of active cell synthesis.

Because nucleoli can fuse with one another, their number is normally variable and cannot be used as a reliable measure of nucleolar activity. In most cells the general trend is for the number of nucleoli to decrease as the cell ages, except for an occasional budding that increases the number of nucleoli; in a few transient types of cell, such as oocytes and nurse cells, the number of nucleoli may increase regularly. The maximum number is fixed. It is determined by the nucleolar organizers; these are differentiated segments of certain chromosomes and therefore occur in fixed numbers under normal conditions [3]. The maximum nucleolar number is therefore an effective index to the degree of ploidy, that is, the number of sets of chromosomes. The maximum number will be found in young cells, before nucleolar fusions have taken place [4].

The exact chemical function of the nucleolus is still not fully understood. However, the fact that the nucleolus and the cell itself are active at the same time was quickly recognized, and it soon became clear that the nucleolus was involved in the cell's functions; eventually it was realized that it was particularly concerned with protein synthesis. During periods of intense protein-production the nucleolus tends to become single and large, rather than multiple and small. This is typical, for example, of gland cells secreting enzyme and mucus protein, of neutrons elaborating Nissl substance (a ribonucleoprotein), and of embryonic and regenerating cells that show a marked production of protein. Again, nucleolar development is correlated with the nutritional level of the cell. However, nucleoli in tumour cells do not show any special characteristics, and although there is a loose correlation between nucleolar appearance and malignancy it is not specific enough to help diagnosis. Nevertheless, some still believe that cancerous growth is initiated by specific stimulation of the nucleolus.

Cells that normally do not actively synthesize protein, such as mature leucocytes and muscle cells, have inconspicuous nucleoli. Similarly, the nucleolus normally disappears during the phase in the cell cycle when the cells are not synthesizing, that is, during cell division (the spindle protein is synthesized before cell division), and cleaving embryonic cells, that are increasing in number without making any protein, have, normally, no nucleoli. Nucleoli reappear in the embryo only at gastrulation, when maternal yolk reserves are exhausted and protein synthesis is resumed. The embryos of mammals, molluscs, and annelids, which have well-developed nucleoli from the earliest stages, are exceptions to this rule. It is not known whether this is because these nucleoli have special functions, although this is suggested by the specialized mode of embryonic development in these phyla, placental in mammals and spiral in the others. There is general, but not complete, correlation between brisk protein-synthesis in cells and their having vigorous nucleoli. Exceptions include protein-deprived liver cells and certain degenerating tumour cells. In the liver cells, the large nucleoli may represent an abortive compensatory development of the protein-synthesis system, or they may indicate, as has been suggested for neurons, a preparation for later demands on synthesis.

The fact that nucleoli are often damaged by agents that depress the cell's metabolic activity suggests that the nucleolus is involved in that activity, as does the fact that pathological conditions of cells are frequently accompanied by nucleolar changes.

It has frequently been claimed that oocytes and cancer cells produce nucleolar substance that is visible under the microscope, but the significance, and even the general occurrence of this is still doubtful, especially as it is rarely observed with the electron microscope. This instrument does, however, clearly show that there are frequently nucleoli on one side of the nuclear membrane, close to endoplasmic reticulum (an important synthetic site in cytoplasm), the Golgi complex, or mitochondria, on the other (figures 3 and 4). This certainly suggests some transfer of nucleolar material, even if only of molecular size. The association of the nucleolus with certain differentiated regions of cytoplasm is sometimes mediated by chromatin condensed around the nucleolus [5]; this was known for many years, and led to hypotheses of nucleolar function that will be discussed later.

THE APPEARANCE OF THE NUCLEOLUS

When treated with dyes, the nucleolus shows certain inhomogeneities, such as vacuoles, inclusions, and filaments. Sometimes these variations correspond to physical artefacts rather than to permanent structures, but at other times they have a real chemical basis and indicate the degree of nucleolar activity. For instance, the formation of large vacuoles is characteristic of highly secretory states, and, sometimes, of degenerative states.

The electron microscope shows the nucleolus as a fine filamentous matrix, presumably proteinaceous, in which numerous denser particles, measuring about 150 Å in diameter, are embedded (figure 4). These are similar to those found elsewhere in the nucleus and in the cytoplasm, and are assumed to consist of ribonucleoprotein. The core of the nucleolus appears denser than the more peripheral region, especially after fixation with osmium; this density may be due either to a denser matrix or to closer packing of particles. This core corresponds to the 'nucleonema' of C. Estable and R. J. Sotelo; the periphery to their 'pars amorpha' [6]. The chromosomal nucleolus organizer often remains embedded in the nucleolus that has grown from it. The appearance of the ultrastructure of the nucleolus depends markedly, as would be expected at this morphological level, on the procedures, such as fixation and embedding, used to study it, but much of the variation undoubtedly reflects physiological changes.

Differentiation of the nucleolus parallels, in certain cells, differentiation of the cell itself. Such cells include oocytes, where nucleoli can be very well developed, and certain special cells of the salivary glands of insect larvae. It has, in fact, been suggested [7] that nucleoli are, on the whole, more concerned with cell differentiation than are chromosomes, and, what is almost equivalent, that the nucleolus is the seat of general cell-heredity (cytoplasmic heredity), while the genes are responsible for the inheritance of more particular traits [8]. This links up with A. M. Dalcq's idea of 'ontomutations', which, by dramatically altering the heredity, would be the source of new major developmental traits [9]. In fact Dalcq has dwelt on the possible relationship between the presence of nucleoli in the early mammalian embryo and its placental development. Some embryos, however, are known to differentiate to some extent, although not to full term, even when they entirely lack organized nucleoli: the implications of this are discussed later.

The nucleolus may best be considered as a differentiated locus of a particular chromosome. This is particularly so in specialized chromosomes, such as the polytene chromosomes of flies; the situation is probably less simple in ordinary chromosomes. There are other chromosomal organelles, such as puffs and Balbiani rings, that, although different in some respects from nucleoli, have several similar physiological and morphological characteristics. Again, there is normally no organized nucleolus in the salivary nuclei of chironomid flies, but, instead, a nucleolar substance spread rather widely on the chromosomes. It is, in fact, possible to speak generally of a nucleolar substance of which the nucleolus is an organized form.

It is not known whether all nucleoli, which presumably have largely similar functions, are similar in origin. This question arises in the case of nucleoli in cells formed during the early division of mammalian eggs, which nucleoli may originate from both chromosomes and other sources. It also arises, especially, in the case of the nucleoli of amphibian oocytes, where some nucleoli are apparently formed as their somatic counterparts in chromosomes, while others are definitely not formed there [10]. Similarity of origin in these cases would be possible only if there were a general nucleolar substance; the evidence appears to suggest that there is.

It is still questionable whether bacteria and and blue algae have nucleoli, but protozoa certainly have them. The organisms without nucleoli are highly successful, by biological criteria; this raises the question, which will be considered later, of the relevance of nucleolar function in the more highly evolved cells.

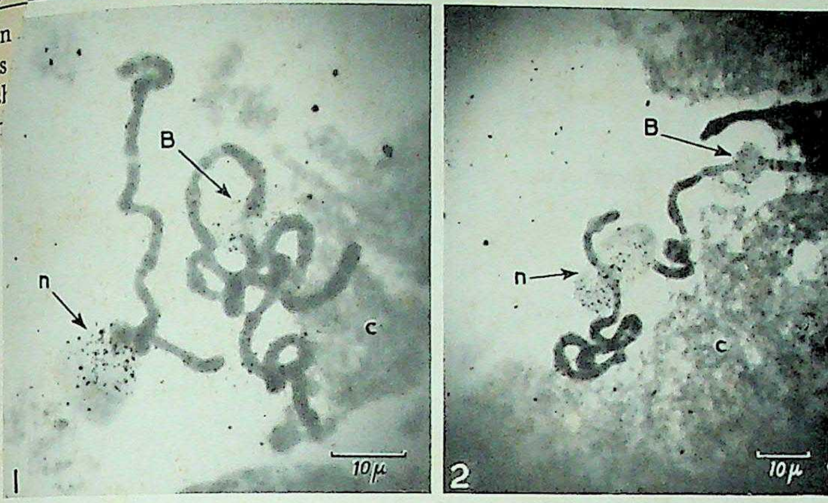
The nucleolus disappears at the end of prophase, before division of the cell, and reappears at telophase after division. Sometimes the nucleolus persists through cell division, being discarded into the cytoplasm, but new nucleoli are formed even in these cases. These persistent nucleoli are more typical of plants than animals; they are associated with certain other morphological characteristics, and so have proved to have a limited taxonomic value. There are a number of possible explanations of the disappearance of the nucleolus. It may occur at the same time as a change in nuclear conditions that is necessary if nucleolar integrity is to be maintained, possibly by making the nucleolus accessible to enzymes that are restricted to cytoplasm during interphase. Another possibility is that nucleolar-chromosomal interactions, for example, demand periodic renewal or reformation of the nucleolus; a third is that partition of nucleoli between the two daughter cells during mitosis proves embarrassing to the cell, which resorts instead to forming new nucleoli. That the nucleolus disappears so as to release nucleolar substance seems unlikely, as new nucleoli are formed even when old ones persist.

The way in which the nucleolus forms anew at telophase offers some insight into the nature, and therefore the significance, of the nucleolar substance. In nuclei with polytene chromosomes, the nucleolus is the product of the chromosomal locus on which it is borne; in ordinary nuclei the nucleolus forms at the organizer, but uses, partly at least, materials formed elsewhere in the nucleus. Microscopy and electron microscopy show, fairly convincingly, 'prenucleolar bodies' in such nuclei, forming amidst the chromosomes and gathering at the organizer to re-form the nucleolus; exceptionally, the organizers are missing and the bodies remain ungathered. The driving force behind the coalescence of the bodies remains obscure, and there is evidence for the control of nucleolus formation by widespread genetic loci as well as by the organizer. There are also several lines of evidence for a competition for nucleolar precursors between different organizers co-existing in a nucleus or, alternatively, for a mutual inhibition between sometimes quite numerous potential organizers in the same nucleus. Thus, when organizers are dislodged into nuclear 'pockets' they form smaller nucleoli, as if they were cut off from precursors by the more centrally located organizers; these form nucleoli of normal size. Again, after irradiation, the lagging chromosomes and the micronuclei derived from them produce a total amount of nucleoli that is greater than normal. The balance of these intra-nuclear interactions favours, in normal circumstances, the predominant organizer, which will form the nucleolus. Swarming of nucleoli, as it occurs in oocytes, pronuclei, and protozoa is thus explainable by a temporary relaxation of competition or inhibition.

THE FUNCTION OF THE NUCLEOLUS

The nucleolus was initially considered to form the nucleus, and that in turn to form the cell [11]. Until recently, it was also thought to give rise to yolk and melanin, and to all the organelles of *Paramecium* [12]. Various other functions proposed have been that the nucleolus is a source of nutritional reserves for the cell in general or for the nucleus in particular; an energy reservoir for the nucleus; a contribution from cytoplasm to nurture the nucleus; a reservoir for materials produced at a limited rate by the chromosomes; and a store for waste from nuclear and cytoplasmic metabolism.

The 'matrix theory' proposes that when the



FIGURES 1 *and* 2 – *Autoradiograms showing the beginning of incorporation of tritiated uridine, a precursor of RNA, in the nucleolus* (*n*) *and Balbiani ring* (*B*) *of salivary gland nuclei of a chironomid. In the nucleolus the incorporation begins close to, but not in, the chromosomal nucleolus organizer at the centre of the nucleolus. Chromosomes and cytoplasm* (*c*) *have not yet begun to incorporate tracer.* (*Reproduced by permission of Butterworths Scientific Publications, London.*)

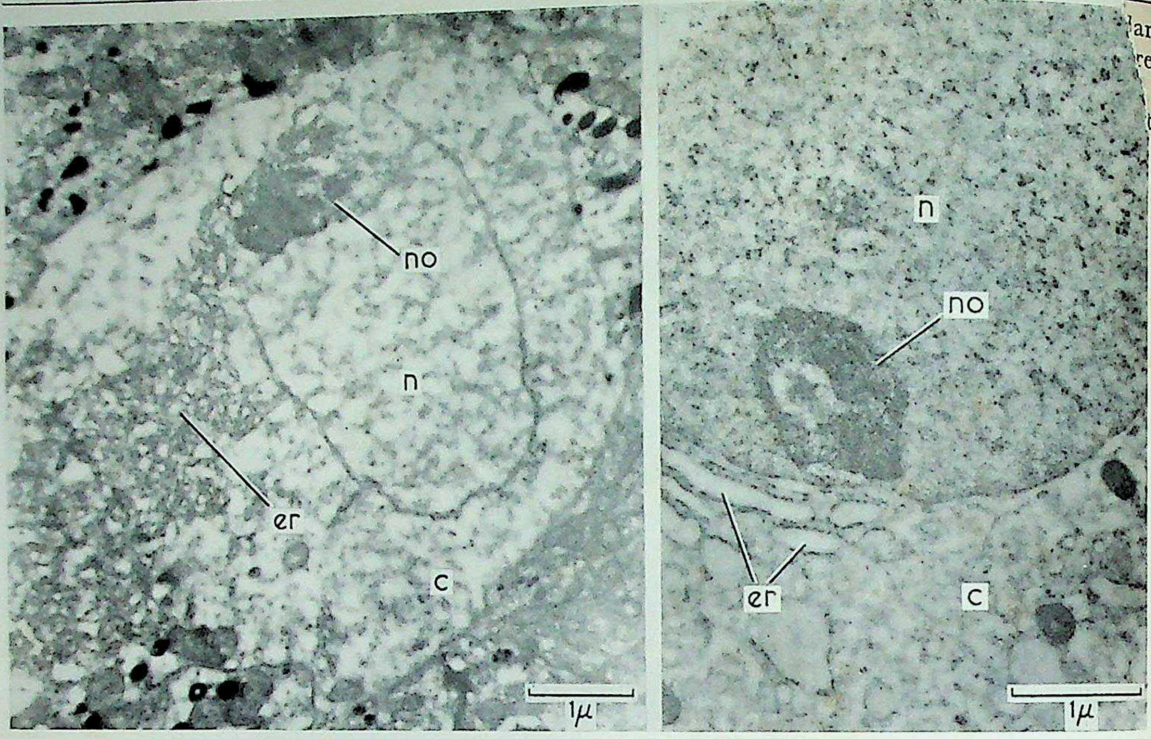
nucleolus disappears at division, nucleolar materials are deposited as a matrix on the chromosomes, and re-enter the new nucleolus formed at telophase. C. D. Darlington [13] speaks of the nucleolus and heterochromatin as alternative sources of nucleic acid for the chromosomes. Certain appearances of the nucleolar apparatus in nerve and gland cells, when it comes into contact with the nuclear membranes through the nucleolus-associated chromatin, suggest that nucleolar substance is discharged into cytoplasm.

Two stimulating modern theories have been put forward by Caspersson and his associates [5], and by Estable and Sotelo [6]. Caspersson considered that the nucleolus-associated chromatin, when stimulated, would produce basic proteins that would gather at the nucleolus and then pass to the cytoplasm, where the synthesis of specific cell-proteins would take place. RNA would be operative in both the nucleolar apparatus and the cytoplasm, the apparatus directing the cell system for protein synthesis. According to this theory, the mechanism gets out of control in cancer cells, and consequently operates with excessive intensity. If the roles attributed in this theory to basic proteins and to nucleolus-associated chromatin are extended to include RNA and chromatin as a whole, the main tenets of this theory remain essentially valid today. Caspersson has recently considered the passage of nuclear RNA to cytoplasm.

Estable and Sotelo regarded the nucleolus as formed of two components of different significance. One, the nucleonema, is deposited on the chromosomes during cell division, and enters the new nucleolus that forms after division; the nucleonema has genetic continuity (*omnis nucleonema e nucleonema*). The second component, the pars amorpha, is a plastic one that disappears at mitosis and forms again at telophase; the pars amorpha would be the metabolic component for each functional phase of the nucleolus during interphase. This conceptually attractive theory, which retains elements of the matrix theory, is more difficult to evaluate from the available evidence than is Caspersson's. There is not enough convincing evidence that the nucleonema is deposited on the dividing chromosomes; such evidence would, of course, be difficult to obtain. What is probably a more serious criticism is that of J. A. Serra [14]. He argues cogently, on the basis of his own observations, for an epigenetic, rather than a genetic, role for the nucleonema.

The discussion that follows deals with the chemical processes that might be performed by the nucleolus; a systematic theory has not been attempted and the ideas are offered to suggest further work. The discussion is based on the assumption that the nucleolus plays a general part in the cells of higher organisms that is indispensable in the long run to cells that produce protein, but essentially auxiliary.

The assumption that the function is general results from the identity of the nucleolar locus in the chromosome (the organizer) in all nuclei of an organism and from the functional equivalence of nucleoli as shown in interspecific hybrid crosses. That it is indispensable for protein synthesis is suggested by the facts that the nucleolus is generally present when the cell is synthesizing protein and that it is an entity, without a membrane, embedded in the genetic code (the cell nucleus); by the weighty circumstantial evidence for its metabolic activity; and by its turnover of RNA, which is known to participate in protein synthesis. Finally, the following facts suggest that the role of the nucleolus is essentially auxiliary. Firstly, embryonic differentiation, which implies synthesis of



FIGURES 3 *and* 4. – *Electron-micrographs of cone cell of larval insect eye (figure* 3) *and of notochord cell of amphibian embryo (figure* 4), *showing the nucleus (n) and the nucleolus (no). The nucleolus appears to be associated with developing endoplasmic reticulum (er) in the cytoplasm (c). (By courtesy of C. H. Waddington and M. M. Perry.)*

specific protein, proceeds to some extent in the absence of organized nucleoli and of organizers. Secondly, cells that do not synthesize protein, and cleaving embryos—both on a 'maintenance metabolism' as it were—have no conspicuous nucleoli, although they still require, for example, maintenance of enzymes. Thirdly, lower organisms, some cells in sciarid larvae, and rapidly dividing tumour cells synthesize proteins, although they seem to have no organized nucleoli. The role of the nucleolus may be the result of its having either one function in cell metabolism or a complex set of functions that acts, for instance, as a rate-limiting system. The nucleolus may also, in certain cases, have specialized functions; these are not considered here. They may result from secondary functional adaptations and need not exclude a more general nucleolar function, which will be discussed in relation to enzymes, RNA, and protein.

ENZYMES

Nucleoside phosphorylase and DPN synthetase, two mainly nuclear enzymes that presumably participate in the synthesis of coenzyme, are localized in the nucleolus [15], and it is therefore possible that the nucleolus is a source of coenzyme for the cell. It has been suggested that when nucleoli and mitochondria occasionally come closer together across the nuclear membrane, the nucleoli provide the mitochondria with coenzyme. Dalcq has proposed that the nucleolus produces adenosine triphosphatase in embryos.

RIBONUCLEIC ACID

The participation of the nucleolus in the RNA turnover of the cell is likely to be its most important function, consisting in making RNA either as the end-product or as an intermediate in the synthesis of some protein. This is one aspect of the wider question of the provision or control by the nucleus of the RNA needed by the cytoplasm for protein synthesis. The dependence of cytoplasmic RNA on that in the nucleus is supported by ample evidence, and the intensity of nucleolar RNA metabolism is in general parallel to that of nuclear RNA. Recent experiments on irradiation of the nucleolus have demonstrated that part of the cytoplasmic RNA is dependent on that of the nucleolus, and to a lesser extent, on chromosomal RNA [16]. This was to be expected from the fact that the nucleolus contains a fair proportion of the total RNA in the nucleus but rarely the whole of it. It is important to note that these experiments also show that a fraction of the cytoplasmic RNA is not dependent on that in the nucleus or nucleolus. This fraction may be synthesized *in situ* in cytoplasm instead of in the nucleus, where the nuclear- or nucleolar-dependent cytoplasmic RNA fractions are supposed to be synthesized. It has recently been shown that, in spider oocytes, the nucleotide composition of nucleolar and cytoplasmic RNA is identical, and different from that of the remaining nuclear RNA; this, again, suggests some correspondence between the two former RNAs [17]. The genetic implications of the nuclear origin of cytoplasmic RNA will become clear later.

The local synthesis during interphase of at least a part of the nucleolar RNA is safely inferred from the existence in the nucleolus of all the precursors and enzymes for RNA synthesis, and from biochemical and autoradiographic studies. Generally, the more differentiated the cell is, the greater the rate of synthesis. During reconstitution of the nucleolus at telophase some of the nucleolar RNA is apparently contributed by the chromosomes in the form of the prenucleolar bodies already described. One possibility to be seriously considered is that of a continuous contribution of chromosomal RNA to that in the nucleolus even during interphase. Nucleolar function during interphase and nucleolar reformation during telophase would then not be different processes but two aspects of the same process. The impression gained from the evidence as a whole is that the turnover of RNA is a property of the nucleolar substance rather than, for instance, of the organizer, irrespective of whether it has coalesced as an organized nucleolus or not. However, the most active site of RNA turnover for different nucleoli varies according to their organization. It is central in the chironomid salivary nucleolus, which remains attached to the central organizer (figures 1 and 2), and it is spread over the surface of the *Drosophila* nucleolus, which is detached from the organizer.

Nucleolar RNA may be related to RNAs with different biochemical functions, transfer, ribosomal, and messenger RNA. Whatever the function of nucleolar RNA may be, the nucleolus could serve *inter alia* as a store for it. This is suggested by its mode of formation, as well as by instances of nucleoli becoming surcharged with RNA as the result of altered physiological conditions. For more details the readers is referred to a review article by the author [18].

Transfer RNA exists in a low-molecular form in the cell sap; its function is to accept the amino acids presented by specific amino-acid-activating ensymes, and to deposit them in ribosomal RNA, where they are synthesized into protein. It is believed that the origins of transfer RNA may be in the break-down of ribosomal RNA, or synthesis in the cell sap or the nucleus.

Synthesis of transfer RNA was found in the two types of nucleolus that have been investigated, namely in the sea-urchin oocyte [19] and in the chironomid salivary cell [18]; it is likely that it occurs in most other nucleoli. Transfer RNA was characterized by biochemical criteria in the oocyte and by the incorporation of radioactive pseudouridine, a fairly specific precursor, in the salivary cell. It was shown that in the latter the nucleolus is an important source of transfer RNA in the cell.

The synthesis of transfer RNA indicates a definite function for the nucleolus in protein synthesis in the cell, either in the supply of this RNA as such, in its use in the nucleolar synthesis of a protein, or for storage of activated amino acid. Transfer RNA does not convey genetic information, but has instead an auxiliary function in protein synthesis that corresponds well with the overall auxiliary function in the cell contemplated for the nucleolus itself. It is, therefore, possible that the nucleolus evolved in adaptation to the function of transfer RNA in protein synthesis. This function could be a primordial one, in an evolutionary sense, or evolved from systems lacking transfer RNA but having alternative pathways, the two systems coexisting today in the same cell. Transfer RNA occurs in lower organisms that do not possess organized nuclei and nucleoli: that these organisms never develop multi-cellular complexity may prove significant.

The high-molecular RNA of ribosomes exists in the ubiquitous ribonucleoprotein particles of the cell that are concerned with protein synthesis. This serves as a template for the conversion of amino acids into protein, the acids being presented by transfer RNA. For this task ribosomal RNA should contain some information for the proper sequential alignment of amino acids. This is in principle coded in the genes, and it is thought that ribosomal RNA is synthesized in the nucleus under their control, and is later conveyed to the cytoplasmic sites of synthesis, where it fulfills its role of template. Since the nucleolus itself represents a gene, ribosomal RNA could also be directly synthesized there. Messenger RNA is assumed to programme the ribosomes, and therefore also to contain some sequence information; it would be synthesized under gene control.

Evaluation of any ultimate nucleolar contribution in ribosomal RNA remains speculative, because of the need for means of chemical characterization *in situ*. The synthesis of transfer RNA in the nucleolus, discussed in the previous section, does not exclude that of ribosomal RNA, as all the four major ribonucleosides that occur in the nucleolus are common to both types of RNA. The electron microscope shows that the ribosomal particles present in the nucleolus are similar to those in the nucleus and cytoplasm. It might be that these particles are synthesized in the nucleolus and are destined to go to the cytoplasm. They could either be completely synthesized there or only the protein moiety; in the latter event, their RNA would presumably be synthesized elsewhere in the nucleus. The biochemical evidence favours a nuclear origin for the complete particle. Alternatively, these particles could be the nucleolar site for synthesis of protein and not be destined for export to cytoplasm. Coexistence of two particle populations in the nucleolus, one being part of the local machinery of synthesis and the other the product of this synthesis, is also possible. Furthermore, it is conceivable that a single particle could display synthetic activity first in the nucleolus and later in the cytoplasm.

Currently, the nucleolus is regarded as in some way assisting in the replication of genic RNA, which otherwise would not be produced in sufficient amounts for the cytoplasmic sites of synthesis, particularly at times of very active protein-synthesis, such as during secretion or differentiation. Genic RNA is copied using master templates that are deposited in the nucleolus. During reconstitution of the nucleolus at telophase, this could be achieved through the prenucleolar bodies. This would imply a flux of nuclear RNA during interphase towards the nucleolus, but the evidence of this from autoradiography is inconclusive. If the ribosomal and messenger RNAs are gene-specific, as assumed, it follows that the nucleolus should incorporate master templates from as many genes as require nucleolar collaboration for the replication of these RNAs. It has also been suggested that nucleolar RNA could bind protein for storage, and *vice versa*.

Three types of template RNA are therefore needed in the nucleolus, to fulfil an auxiliary role in the system of protein synthesis:

1. A single template (or at most a few templates) for production of a general nucleolar RNA, transfer or ribosomal RNA. Since the nucleolus itself represents one gene (or at most a few), this template can be nucleolar RNA proper and not RNA from the chromosomes. There is, in fact, evidence that in polytene chromosomes nucleolar RNA is the product of the nucleolus itself.

2. A template for production of a general nucleolar end-product, such as ribosomal protein. This protein can, as in the case of transfer or ribosomal RNA, be produced by one or a few templates consisting of nucleolar RNA proper; it is considered later.

3. Template RNAs of chromosomal origin gathered at the nucleolus for replication of ribosomal or messenger RNA. Even if some genes require the assistance of the nucleolus to produce these RNAs, it is conceivable that not all will.

The chromosomes most likely to be self-sufficient are the multiple-stranded polytene chromosomes of diptera, and there is evidence for widespread liberation of RNA by them. (If all genes were self-sufficient, the template RNA of type 3 would be unnecessary.) Support for these views comes from the fact that polytene chromosomes of sciarid diptera normally have no nucleoli but have instead a widespread nucleolar substance, and that certain mammalian cells in tissue culture, having ordinary chromosomes, but with smaller nucleoli, also have a widespread nucleolar substance. A fundamental difference between polytene and ordinary chromosomes could thus result from the degree to which the two types require the collaboration of the nucleolus for the replication of template RNAs.

PROTEINS

Although nucleoli of both growing and mature cells possess all the prerequisites for protein synthesis, the turnover of amino acid occurs only in growing nucleoli. By contrast, RNA turns over continuously throughout the life of the nucleolus. This limited turnover of amino acid can be accepted, on various grounds, as reflecting net protein synthesis and not merely amino acid activation or exchange. Some of this synthesis would be of growing nucleolar protein itself and some of it could perhaps be of exportable protein, particularly since it has recently been found that the nucleolus utilizes, at least partly, pre-existing protein for its growth. It is clear that the fully mature nucleolus can exert control over cell proteins only through proteins previously synthesized during its growing period, or through the agency of nucleolar

RNA that continues to turn over during nucleolar maturity.

Three classes of exportable nucleolar protein will be considered. The template mechanism for producing these proteins would be as postulated for RNA in the previous section.

Several lines of evidence suggest [18] the possibility of a nucleolar production of ribosomal protein. This could combine with the ribosomal RNA to produce the finished ribosomal particle *in situ*. This is, in fact, one of the simplest overall functions that can be proposed for the nucleolus, having proper regard to its auxiliary role in the cell system for protein synthesis, which would account for all that is known of nucleolar turnover of amino acid and RNA. It implies that ribosomal particles are produced by the nucleolus only while it grows.

Several other proteins can be mentioned here. First, in some oocytes there is in both nucleoli and yolk a common antigen, which could have a nucleolar origin; it has long been supposed that the nucleolus plays a part in yolk formation. In *Acetabularia*, nucleolar proteins have been implicated in the production of morphogenetically active proteins in cytoplasm. Both these proteins could in fact be ribosomal.

Nucleolar emission of protein, sometimes as 'matrix', from the nucleus is common, if not general. It was suggested that antibody and hormone proteins are accumulated in nucleoli of the cells in which these substances are metabolized. It has also been claimed that haemoglobin is synthesized in the erythroblast by the nucleolus [20].

Several similarities between nucleoli and spindles suggest that the first might be involved in the formation of the second. Firstly, both appear to contain only one protein, as well as having other gross similarities of chemical composition. Secondly, there is an apparent association of the nucleolus with spindle function [18]. Thirdly, formation of the spindle is rapidly inhibited by irradiation of any single nucleolus in a nucleus. Fourthly, spindle protein is synthesized before spindle formation, that is before the nucleolus has disappeared at cell division and perhaps while it is still turning over amino acid. The relatively large amount of spindle protein in the cell shows, however, that not all spindle protein is likely to be directly dependent on nucleolar synthesis. An alternative proposal is that the nucleolus mediates the release of agents that are structurally active with respect to the spindle. The most obvious argument against the proposition that the only function of the nucleolus is connected with spindle formation is its manifest activity in cells that secrete but do not divide.

REFERENCES

[1] HUGHES, A. 'The Mitotic Cycle'. Butterworths Scientific Publications, London. 1952.

[2] MONTGOMERY, T. H. *J. Morphol.*, **15**, 265, 1898.

[3] HEITZ, E. *Planta*, **12**, 775, 1931.

[4] SCHREIBER, G. *Z. mikr.-anat. Forsch.*, **66**, 301, 1960.

[5] CASPERSSON, T. O. 'Cell Growth and Cell Function'. W. W. Norton, New York. 1950.

[6] ESTABLE, C. and SOTELO, R. J. *Inst. Inv. Cien. Biol. Publ.*, **1**, 105, 1951.

[7] WILLMER, E. N. 'Cytology and Evolution'. Academic Press, New York. 1960.

[8] BERNHARD, W. and OBERLING, C. 'Problèmes d'ultrastructure et de fonctions nucléaires' (edited by Thomas, J. A.), p. 279. Masson et Cie., Paris. 1955.

[9] DALCQ, A. M. *Bull. Soc. zool. Fr.*, **79**, 240, 1954.

[10] CALLAN, H. G. and LLOYD, L. *Phil. Trans.*, *B*, **243**, 135, 1960.

[11] SCHLEIDEN, M. J. 'Contributions to Phytogenesis' (appendix to SCHWANN, T. 'Microscopical Researches into the Accordance in the Structure and Growth of Animals and Plants'). Sydenham Society, London. 1847.

[12] EHRET, C. F. 'Symposium on Information Theory in Biology' (edited by Yockey, H. P.), p. 218. Pergamon Press, New York. 1958.

[13] DARLINGTON, C. D. *Symp. Soc. exp. Biol.*, **1**, 252, 1947.

[14] SERRA, J. A. *Rev. Port. zool. biol. gen.*, **2**, 117, 1959.

[15] BALTUS, E. *Biochim. Biophys. Acta*, **15**, 263, 1954.

[16] PERRY, R. P. *Exp. Cell. Res.*, **20**, 216, 1960.

[17] EDSTRÖM, J. E. *J. Biophys. Biochem. Cytol.*, **8**, 47, 1960.

[18] SIRLIN, J. L. *Progr. Biophys. biophys. Chem.*, **12**. (In press.)

[19] VINCENT, W. S. and BALTUS, E. *Biol. Bull.*, **119**, 299, 1960.

[20] O'BRIEN, B. R. A. *Exp. Cell. Res.*, **21**, 226, 1960.

Reactive dyestuffs for cellulose

I. D. RATTEE

Dyeing is one of the most ancient of technologies, yet until little more than a hundred years ago the preparation and use of dyes remained more of an art than a science. The last century has seen not only the discovery of a vast range of synthetic dyestuffs but a great increase in our understanding of the dyeing process. One recent result of this has been the development of a new and important class of dyestuffs—the reactive dyes, so called because they react chemically with hydroxyl groups of the fibres being treated. The ideal reactive dye has yet to be found, but the immediate success of this new class has greatly stimulated research.

Until little over a century ago, dyeing was carried out on small batches of material, using empirical methods of applying extracts of plants, woods, and insects. The expansion of textile production that was so significant a feature of the industrial revolution resulted in the rationalization of textile dyeing and its establishment as an advanced industrial technique. The discovery of the first synthetic dye in 1856 by W. H. Perkin, and the subsequent discovery by P. Griess of the diazo reaction, so important in the development of synthetic dyes, provided a basis for the very rapid growth of dye manufacture. The years following these early discoveries saw the preparation of a vast number of synthetic dyes, and today over 3500 separate chemicals, of which 2000 contain an azo group, are marketed as dyes. The volume of research involved is indicated by the fact that these represent a selection from about one million speculative preparations made in industrial research laboratories. G. S. J. White [1] estimated that in 1958 world expenditure on dyes was about £300 million per annum, and this tremendous quantity of colour is consumed by dyehouses, which in some cases produce as much as ten million yards of dyed cloth a week.

Certain types of organic chemical structure have been found to be of special significance in dye chemistry. For example, substances containing two or more azo links have been found to be very valuable for cheaply dyeing cellulosic fibres such as cotton and rayon, though with only moderate fastness. Where a particular structure has been found to be of special value, groups of dyes that can be applied in the same way and that exhibit the same general resistance to wet treatments have been developed. It is common practice to refer to direct cotton dyes, vat dyes, azoic dyes, and so on, thereby denoting dyes having similar general properties.

The importance of cellulose fibres as textiles has resulted in the development of some eight classes of dyes for them. The most recent addition to these has been the reactive dyes; these have developed more rapidly in use, and aroused more interest, than any previously discovered. Despite the fact that it is still a relatively small group, numbering just over a hundred named products, it is the subject of more research and patent activity than any other. There are two main reasons for this, which are best explained by a general consideration of the dyeing process itself.

When a textile is immersed in a solution of dye, the solution may be merely imbibed or the dye may transfer from the solution to the fibres. Dyeing properties are conferred by the presence in the dye molecule of certain chemical arrangements or groups that encourage either solution in the fibre or the formation of hydrogen bonds or electrovalent links with it. All classes of dyes are absorbed on all fibres by this basic mechanism. The second important property of a dye is its mobility in the fibre; this determines the resistance of the dyeing to washing or other wet treatments. This resistance may be influenced by the formation of hydrogen bonds, electrovalent links, aggregation of the dye in the fibre, or in other ways. In general, dyes that are held only by weak links, such as hydrogen bonds, do not show a high degree of wet fastness, because of the inherent reversibility of the mode of attachment.

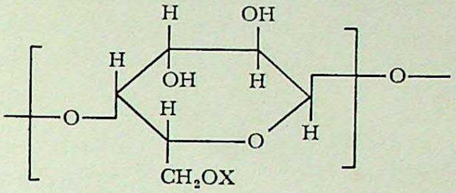
To produce a high degree of wet fastness, it is necessary to bring about some change in the nature of the dye after it has been absorbed by the fibre, so that it differs from its original form. One way is to make the dye insoluble. For example, vat dyes are pigments containing quinone systems, which are first made soluble by alkaline reduction to their sodium salts. After absorption on the fibre, the reduced dye is oxidized to form the insoluble pigment, which is retained mechanically and is highly resistant to washing. Alternatively, the first stage may involve the use of a substance which is not itself a dye but is made to react with another agent, such as a diazonium salt, to form a pigment in the fibre. This last method has been highly developed; it is now even possible to synthesize copper phthalocyanine derivatives on the fibre in this way [2].

Another possible method of permanently retaining the dye in the fibre, but one only recently achieved, is to use a compound that contains labile atoms or groups able to react with the fibre substance itself through a covalent bond; this is the basis of the reactive dyes. Before their advent, it was possible to produce fast dyeings on cellulosic fibres only by means of one of the 'pigmentation' techniques outlined above. To produce a dye that will work in this way, it is necessary to incorporate into the molecule particular groups, such as the quinone group, and this severely limits the dye chemist in developing new shades. In addition, making the dye insoluble results in duller tones. The reactive dyes attach themselves to the fibre by a group that can be separate from the chromophoric system; consequently a very much wider range of shades, including brilliant shades previously unobtainable on cellulosic fibres with any useful degree of fastness, can be obtained.

The production of fast dyeings on cellulosic fibres by a 'pigmentation' technique involves treatment after the initial dyeing process, possibly with several stages. For example, with vat dyes the sequence of operations is reduction→dyeing→oxidation→soaping. The application of reactive dyes is very much simpler, and involves, at most, two stages.

The popularity of reactive dyes thus springs from two characteristics, and their clear technical advantages have caused the first reactive dyes for cellulose—introduced in 1956 by Imperial Chemical Industries under the trade name Procion—to be joined rapidly by other products such as Cibacron (Ciba), Remazol (Hoechst), Drimarene (Sandoz), and Reactone (Geigy). Although the full extent of trade reaction cannot have been foreseen by early workers in the field of coloured cellulose derivatives, some of the potential advantages of this mode of dyeing have long been realized, and the commercial development of reactive dyes for cellulose must be regarded as being, in a sense, the culmination of researches that have been going on for sixty years or more.

Work on the production of coloured cellulose derivatives has been confined to esterification or etherification reactions, generally at the hydroxyl group attached to the 6-carbon in the anhydroglucose unit

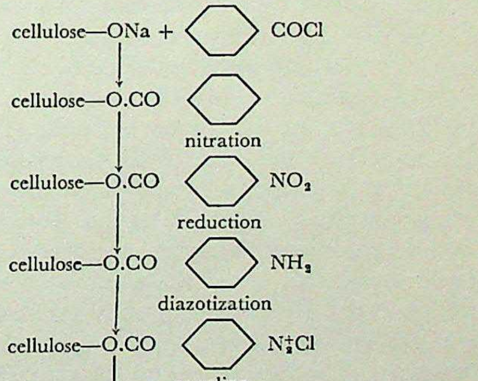


where X is an acyl or aroyl group in the esters and an alkyl or aryl group in the ethers.

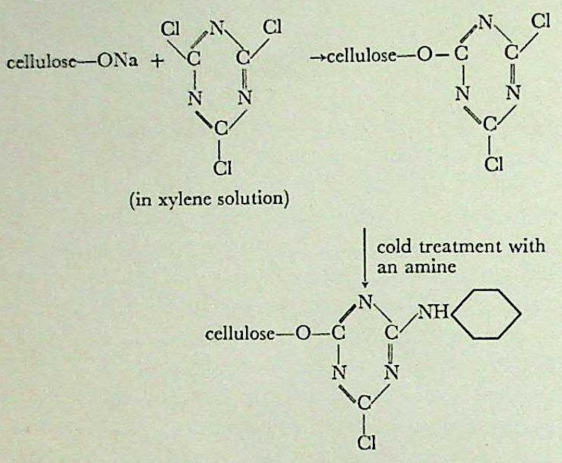
Of the two, the esters of cellulose are the more easily formed; their stability varies with the strength of the esterifying acid, the stronger acids giving the less stable esters. The ethers are in general much more difficult to form, but are more stable. This is not an entirely clear-cut difference, however, for the stability of the product depends on the nature of the ether or ester formed. Because of their relative ease of preparation, attention has naturally tended to be concentrated on the esters.

EARLY RESEARCHES

Among the earliest workers in this field were C. F. Cross and E. J. Bevan [3], who treated cellulose with strong caustic soda solution and then carried out the following sequence of reactions on the 'soda cellulose' so formed:

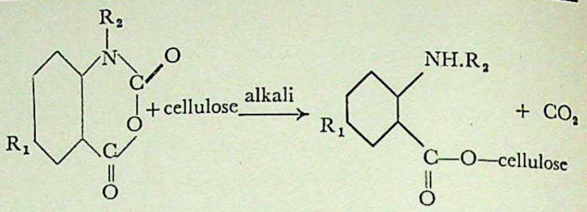


Although the colour of the cellulose derivative so produced was very fast to washing, the nature of the textile was so much altered that it was of little practical value. This approach was very thoroughly explored by many other workers [4–7] using alternative acylating agents, with the aim of shortening the process. From the point of view of subsequent developments, the investigations of Haller and others [8] in 1930–33 are of interest. In this work, the use of 2:4:6-trichloro-1:3:5-triazine for acylating cellulose was examined, using the well-known stepwise reaction-behaviour of this compound; this enables the three chlorine atoms to be reacted separately in the following sequence:



Coloured cellulose may be obtained from the last of these derivatives by coupling with a diazonium salt. However, the reaction procedure used did not depart fundamentally from that of the earlier work. It involved the use of organic solvents, and gave a considerably degraded cellulose of little value for textile purposes.

The dominant feature of all the early work was the use of very strong caustic soda solutions to prepare soda cellulose; it was assumed that cellulose was relatively inert and required vigorous conditions to produce an adequate reaction. This seemed a logical consequence of the work on the preparation of cellulose derivatives for textile and other purposes, where it is necessary to achieve a reasonably high degree of substitution in the cellulose chain. From the point of view of dyeing, however, the degree of substitution required is rarely as much as 1–2 per cent of the total number of hydroxyl groups present, and once this was realized a practicable process became reasonably possible. The possibility was confirmed by F. Günther [9], who was able to esterify cellulose under relatively mild conditions, as shown at the top of the next column.

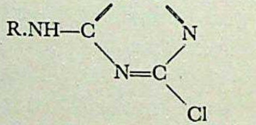


DEVELOPMENT OF PRACTICAL ESTERIFYING DYES

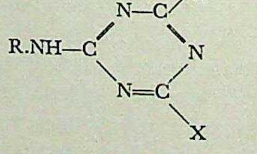
However, until the 1950s there was no extension of Günther's work to cover the use of more satisfactory esterifying agents. I. D. Rattee and W. E. Stephen then examined the use of 2:4-dichloro-1:3:5-triazin-6-yl-amino dyestuffs for the production of coloured cellulose derivatives under the mildly alkaline conditions that are suitable for a dyeing process. This work showed that, if the dyes were suitably forced on to the fibre from the aqueous solution, they were then fixed rapidly. Although the degree of substitution was not high, the efficiency of the reaction was sufficient to provide the basis for attractive dyeing processes yielding fast dyeings over a satisfactory range of shades and depths of shade [10].

It was shown that reactions between these dyes and cellulose could be complete in as little as five seconds, using alkali no stronger than sodium carbonate. Furthermore, it was found that, by varying the pH or temperature, the reaction rate could be controlled so that the dyes could be applied under all the conditions encountered in normal dyeing practice.

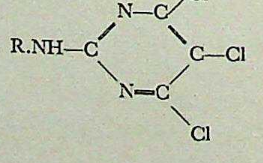
Following these observations, the first commercial range of reactive dyes for cellulose was developed under the trade name Procion. Their success led to intensive industrial research on dyes bearing potential esterifying groups, and this has led to the development of the three following types of reactive esterifying dyes for cellulose:

(1)


Procion dyes (I.C.I.)
In these dyes both chlorine atoms are labile.

(2)


Procion H dyes (I.C.I.)
Cibacron dyes (Ciba)
In these dyes the group X is non-labile.

(3)


Reactone dyes (Geigy)
Drimarene dyes (Sandoz)
In these dyes only one chlorine atom is labile.

COLOURED CELLULOSE ETHERS

Studies of the formation of cellulose ethers resulted in the development of another type of reactive dye which reacts with cellulose fibres under the same general conditions as the esterifying dyes do. These Remazol dyes (Hoechst) [11] are of the general form $R.SO_2.CH_2.CH_2.OSO_3Na$. They form cellulose ethers by elimination of sulphuric acid through the formation of the extremely active etherifying agent $R.SO_2.CH{=}CH_2$.

Coloured ethers have received rather less attention than have the corresponding esters, no doubt because of the relative difficulty of forming them before the discovery of the use of the vinyl sulphonyl group for this purpose. In 1926, D. H. Peacock [12] prepared a fast-dyed cellulose by boiling cotton in an aqueous solution of m-nitrobenzyl-dimethylphenyl ammonium chloride, followed by reduction, diazotization, and coupling. Fibres that could be diazotized were also prepared by C. Granacher [13] by superficial etherification of anhydrous alkali cellulose, using p-nitrobenzyl chloride and similar chloromethyl compounds, or using dinitrohalogenobenzenes with subsequent reduction. Similar experiments were carried out by other workers [6, 14, 15, 16]. The first direct use of dyes themselves for etherifying was made by J. D. Guthrie [17], who found that fast dyeings could be produced by impregnating cotton with sulphate esters of alcohols in caustic soda solution and baking at 100–110° C for an hour or more.

As with the esters, none of this early work led to a practical dyeing process. Only with development of the sulphato ethyl sulphonyl dyes in 1944 [11, 18] and their subsequent application to cellulose, using the reaction conditions previously established in the cellulose ester field [10], was success achieved.

REACTION KINETICS

The four reactive-dye systems currently available are similar in that they all involve reaction with cellulose hydroxyl groups and with water under alkaline conditions, and that in each case a single reaction is decisive in determining the reaction kinetics. This is true of the behaviour of the potentially bifunctional 2:4-dichloro-1:3:5-triazin-6-yl-amino dyes, as well as of the other dyes that are clearly mono-functional, owing to the stepwise character of the reactions of dihalogenotriazinyl compounds with amino or hydroxy compounds. The reaction with water of one of the chlorine atoms in these dyes gives rise to the relatively unreactive monochloromonohydroxy compound, so that as far as the kinetics of dyeing is concerned, the reactions of only one chlorine atom need be considered in the first instance. For this reason, the study of the kinetics of the reactions between dichlorotriazinyl dyes and water or cellulose, to which attention is now turned, provides information that is valid for all reactive systems at present available. Indeed, it can be argued that the theory will be valid for any reactive dye stuff applied to any fibre, provided that only reactions with fibre and with water are involved. The theory is therefore of very wide importance.

In examining this system, it is first necessary to consider the side-reaction with water, for the extent to which this occurs during dyeing obviously governs the efficiency of the process. Further, as will be demonstrated later, the rate of reaction with water at a given *p*H indirectly measures the reactivity of the dye with cellulose.

The rate of hydrolysis of a dichloro-s-triazinyl dye in water can be measured by a number of methods, but since the hydrolysis of the first chlorine atom only is under consideration, a colour reaction between dichloro-s-triazinyl dyes and pyridine in the presence of caustic soda provides an elegant method of analysis [19, 20]. Using this technique, the reaction constants for a number of dichloro-s-triazinyl dyes with water have been determined [21]. A wide variation in reaction rate has been demonstrated, and it has been shown that the reaction rate depends on *p*H.

At *p*H 10, the reactivity with water of the most reactive dye among those commercially available is eighteen times that of the least reactive. To determine whether this is true also of cellulose presents difficulties, and for this reason initial experiments on the reaction of the dyes with hydroxylated substances were carried out with model compounds, such as sorbitol,

$$CH_2OH(CHOH)_4CH_2OH.$$

The sorbitol–dye reaction product itself gives a colour reaction with pyridine in caustic soda, so that the course of the reaction must be followed electrometrically. By maintaining the *p*H at 10·5 or less, at room temperature, and using excess sorbitol, pseudo-first-order reaction constants have been determined; the values are shown in table 1. The wide spread of reactivity with both water and sorbitol can be seen, but the relative constancy of the ratio of reaction constants is also evident.

TABLE I

Reaction constants at p*H* 10 *at* 25° *C*

Dichloro-s-triazinyl dye	$k_{H_2O}\,\mathrm{min}^{-1}$	$k_{sorbitol}$ min^{-1}	$k_{sorbitol}/k_{H_2O}$
Procion Yellow R	$6\cdot0\times10^{-4}$	$2\cdot8\times10^{-2}$	47
Procion Brilliant Yellow 6G ..	$3\cdot4\times10^{-3}$	$1\cdot4\times10^{-2}$	41
Procion Brilliant Red 2B.. ..	$1\cdot1\times10^{-2}$	$4\cdot6\times10^{-1}$	42
Procion Brilliant Red 5B.. ..	$4\cdot6\times10^{-3}$	$1\cdot8\times10^{-1}$	39
Procion Blue 3G	$3\cdot7\times10^{-3}$	$1\cdot4\times10^{-1}$	38
Procion Brilliant Blue R	$1\cdot1\times10^{-3}$	$5\cdot1\times10^{-2}$	46
Procion Scarlet G	$6\cdot1\times10^{-4}$	$2\cdot9\times10^{-2}$	44
Procion Brilliant Orange G ..	$8\cdot4\times10^{-4}$	$3\cdot6\times10^{-2}$	43

The mannitol/water system has also been examined over a wide range of alkalinity, using a single dye. The pseudo-first-order reaction constants for mannitol and for water have been measured, and it has been shown that the ratio $k_{mannitol}/k_{water}$ is reduced as the alkalinity is increased. However, this is not necessarily a complete picture of the true situation, since mannitol ions can be formed under the *p*H conditions of the experiment. Reaction with these ions is possible, and that it does occur is strongly indicated by the fact that when the bimolecular reaction constants are calculated on the basis of mannitol O^- and OH^- concentrations, using known *p*K values, the ratio of the two reaction constants proves to be constant at all *p*H values, as shown for a typical dye in table II.

TABLE II

Comparison of pseudo-first-order and bimolecular reaction-constant ratios (mannitol/water system)

[NaOH]	$k_{mannitol}/k_{water}$	$k_{mannitol\,O^-}/k_{OH^-}$
0·1N	1·05	13·5
0·2	0·89	13·5
0·4	0·66	14·3
0·6	0·58	14·7
0·8	0·40	12·2
1·0	0·41	14·5
2·0	0·20	15·8

Mean value $k_{mannitol\,O^-}/k_{OH^-} = 14\cdot1 \pm 1\cdot2$

This result shows that the reaction is governed by the concentration of carbohydrate ion present, and that for a given carbohydrate, the various dyes all show, despite their wide variation in reactivity, the same ratio of reactivity with the carbohydrate ion as compared with the hydroxyl ion.

Attention may now be turned to the reaction with cellulose itself. The heterogeneous character of the cellulose/water/dye system complicates the application of the above reaction mechanism to the dyeing of cellulose, because calculation of concentrations of the various reactants must involve consideration of affinity and diffusion, and because there may be steric factors in the solid phase to be considered. In addition, it is necessary to demonstrate that cellulose ions can exist under given reaction conditions. Fortunately, this kind of problem has arisen in other fields of work, and P. V. Danckwerts [25] has evolved an equation that may be simplified to a form suitable for dyeing phenomena. His equation is

$$Q=c\left(t+\frac{1}{2k}\right)\sqrt{(dk)}$$

where Q = the amount of dye diffusing through unit area of solid surface
c = the equilibrium surface concentration
d = the diffusion coefficient
k = the reaction constant for cellulose
t = the duration of dyeing

This equation has been applied by H. Sumner [22] to the behaviour of reactive dyes on cellulose. As the Danckwerts equation refers to the ideal case of diffusion into a slab of infinite thickness, it was necessary to carry out experiments with viscose film rather than fibres. By determination of the values of Q, c, and d, Sumner was able to calculate values of k over a range of *p*H and to compare these with the reaction rates with water. As in the mannitol/water system, the ratios are not constant and vary between 25 and 10. However, the behaviour of the mannitol/water system suggests that it is in fact the values of k based on reaction with ionized hydroxyl groups that ought to be compared.

Sumner [23] has given strong evidence, based on a Donnan equilibrium and S. M. Neale's value for the *p*K of cellulose [24], supporting the theory that sodium hydroxide is absorbed by a neutralization mechanism. Excellent agreement has been shown between observed values of caustic soda uptake and values predicted on the basis of the neutralization theory. When the reaction constants determined by the use of the simplified

Danckwerts equation are adjusted to allow for [cellulose O⁻], and the ratios of the new reaction constants with those for [OH⁻] are calculated, the latter are found to be almost constant at all alkaline *p*H values (table III).

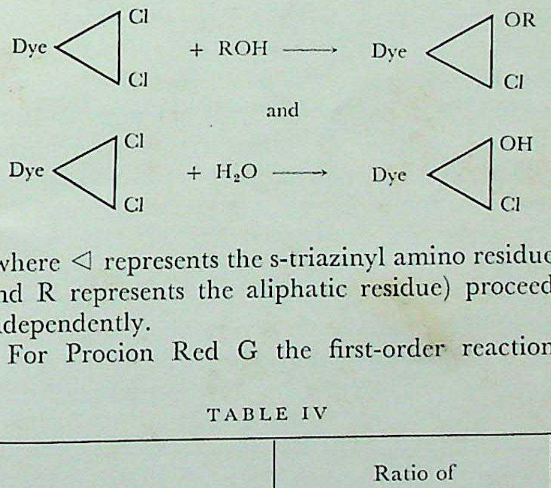
TABLE III

Comparison of pseudo-first-order and bimolecular reaction-constant ratios (cellulose/water system at 20° C)

*p*H	k_{H_2O} min⁻¹	$k_{\text{cellulose}}$ min⁻¹	$k_{\text{cellulose}}/k_{H_2O}$	$k_{\text{cellulose O}^-}/k_{[OH]^-}$
8·00	$9{\cdot}12 \times 10^{-4}$	—	—	—
8·81	$3{\cdot}63 \times 10^{-3}$	—	—	—
9·14	$5{\cdot}62 \times 10^{-3}$	—	—	—
9·46	$6{\cdot}92 \times 10^{-3}$	$1{\cdot}72 \times 10^{-1}$	25	0·95
9·74	$1{\cdot}29 \times 10^{-2}$	$3{\cdot}30 \times 10^{-1}$	26	1·16
10·99	$2{\cdot}82 \times 10^{-2}$	$7{\cdot}03 \times 10^{-1}$	25	1·16
11·70	$6{\cdot}46 \times 10^{-2}$	1·59	25	1·28
11·94	$1{\cdot}00 \times 10^{-1}$	1·87	19	1·06
12·63	$4{\cdot}90 \times 10^{-1}$	7·71	16	1·37
12·73	$6{\cdot}03 \times 10^{-1}$	6·52	11	1·03
		Mean value = 1·14 ± 0·14		

From these experiments an extremely simple picture emerges of the reaction between reactive dyes and cellulose. Reaction takes place with hydroxyl ions and with ionized cellulose, and since the concentration of each of these is extremely small in the absence of alkali, the rate of reaction is negligible. The addition of alkali increases the rate of both reactions, but the preferential reaction with the cellulose cannot be explained on the basis of an inherently greater reactivity of cellulose, because both reaction constants (that is, for cellulose O⁻ and for OH⁻) are of the same order of magnitude. Other factors must therefore be considered. Firstly, there is the affinity of the dye for the cellulose fibre, which causes the dye concentration in the fibre to be as much as 500 times greater than that in the water. Secondly, there is the lower dissociation constant for cellulose as compared with water (*p*K 13·7 as against *p*K 15·7), giving, at any given *p*H, a greater degree of ionization in the cellulose relative to the water. These two factors are sufficient to outweigh the considerable preponderance of water in the practical dyebath. The Danckwerts equation suggests that the affinity of the dye for the fibre, expressing itself through the value of *c*, is the most decisive factor, because the other influences, diffusion and reactivity, are expressed as square roots in the equation. However, high affinity is frequently associated with low diffusion, and some balance between the two is necessary in a useful dye.

The reaction constants determined by Sumner for the reaction of cellulose with the reactive dye are unlikely to refer to a single dye-fibre reaction. T. S. Gardner and C. B. Purves [26] have shown that toluene-p-sulphonyl chloride reacts with the three hydroxyl groups of cellulose at C_6, C_2, and C_3 in the ratio 23·4 : 2·16 : 0·106. There is therefore the strong probability that in Sumner's work the esterifying reaction involved both primary and secondary hydroxyl groups. Since all three possible reactions are of the same order, this will be concealed in $k_{\text{cellulose}}$ determinations and will in no way invalidate the results. Because it is not possible to determine the actual sites in the cellulose molecule at which reactive dyes attach themselves, this has been studied indirectly, using dichlorotriazinyl amino dyes and model compounds. The rates of reaction of a number of such dyes with n-propanol and isopropanol have been determined [19]. Because the dyes are insoluble in the pure alcohols, the experiments were carried out in 50 per cent solutions of the alcohol in water, and a correction for the reaction rate in water alone was applied. This device assumed that the reactions



(where ◁ represents the s-triazinyl amino residue and R represents the aliphatic residue) proceed independently.

For Procion Red G the first-order reaction

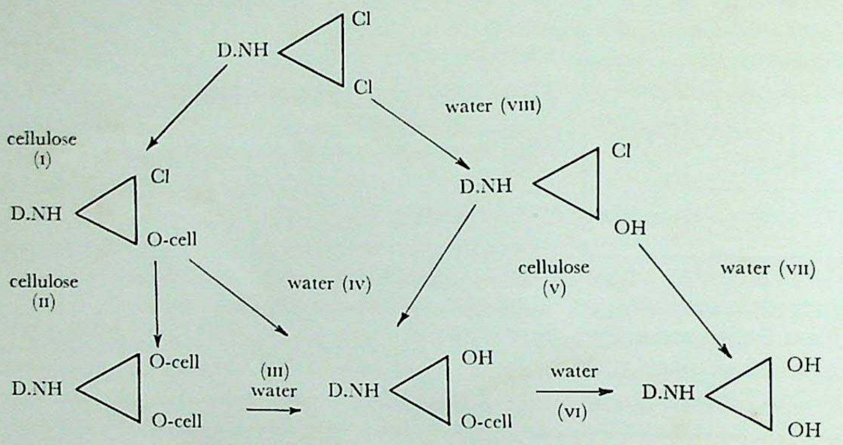
TABLE IV

Dye	Ratio of reaction constants (primary/secondary)
Procion Brilliant Red 2B	3·7
Procion Brilliant Blue R	6·6
Procion Red G	7·2
Procion Yellow R . .	7·2
Procion Blue HB* . .	13·4
Procion Brilliant Red H3B*	15·2

* Monochlorotriazinyl dyes.

constants (k(min^{-1})) were found to be 0·58 for n-propanol and 0·08 for isopropanol. The ratio of the two reaction constants is thus in moderate agreement with the ratios observed by Gardner and Purves. It is to be expected that the ratio will vary with different acylating agents, and this is well demonstrated by the following comparison (table IV) of the reaction-rate ratios for different dichlorotriazinyl dyes.

Reactive dyes bearing other reactive groups behave similarly. These are, however, different from the dichlorotriazinyl dyes in that they are not only less reactive but also possess a single reactive centre. The dichlorotriazinyl dyes, on the other hand, can pursue a multiplicity of reactions, as shown below,



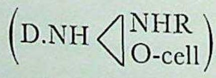
where D is the dyestuff residue and ◁ is the s-triazine residue. Reaction I has been discussed; reaction V is very slow and may be virtually ignored; and reaction VI does not occur under conditions encountered in dyestuff application. Thus the rate-determining reactions are I and VIII, and the picture of the reaction kinetics previously drawn remains true. Reactions II, III, and IV are also of interest, however, because they give rise to dye-cellulose reaction products having different properties.

TABLE V

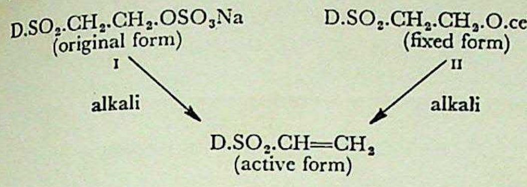
	Acid hydrolysis	Alkaline hydrolysis
Type 1 ..	Relatively stable, slowly converts to type 3.	Initially converts to type 2 and ultimately to type 3.
Type 2 ..	Relatively stable	Relatively stable, ultimately converts to type 3.
Type 3 ..	Relatively unstable	Relatively stable

Cross-linking (reaction II) and subsequent cross-link rupture have been demonstrated by numerous workers [27–29]. In practical dyeing three types of dye-cellulose product may be formed: these may be designated type 1, type 2, and type 3, arising from reactions I, II, and III or IV respectively. Because of the stepwise reactivity of the halogens on halogeno-s-triazinyl compounds, it is possible to produce dyeings that are almost perfect examples of the three types, and the properties of these with respect to acid or alkaline hydrolysis have been studied [21]. These are summarized in table V.

LIMITATIONS OF REACTIVE DYES

The above results have practical significance, because the fastness of some dichlorotriazinyl dyes changes during storage of dyed materials. For example, type 3 dyeings of a few dichlorotriazinyl dyes gradually lose their high initial fastness, and fabrics dyed with them 'bleed' when put into water; this is because the dye-fibre bond ruptures. As types 1 and 2 of the same dyes change slowly to type 3 during storage, and more rapidly if treated with alkali, they too deteriorate, though not so quickly. This effect is offset industrially in two ways. Firstly, since not all dyes show the effect, it can be minimized by careful selection during the initial investigation. Secondly, use can be made of the remaining reactive halogen in type 1 dyeings by causing it to react with an amine to produce what is known as a type 4 dyeing  $\left(\mathrm{D.NH}\triangleleft\begin{matrix}\mathrm{NHR}\\ \mathrm{O\text{-}cell}\end{matrix}\right)$, which is relatively stable.

This problem is encountered to a greater or lesser extent with all acylating dyes, that is, those based on s-triazine or 1 : 3-diazine. The dyes that form vinyl sulphone are free from this trouble, but possess defects of their own. The compound $\mathrm{D.SO_2.CH_2.CH_2.O}$-cellulose formed by the reaction of these dyes with cellulose is related to their original form, $\mathrm{D.SO_2.CH_2.CH_2.OSO_3Na}$, if cellulose is regarded as a weak acid. Under alkaline conditions causing ionization of cellulose, the dye-fibre compound tends to revert to the vinyl sulphone, as shown overleaf.



Cellulose, being a weaker acid than sulphuric acid, requires a higher *p*H to bring about this reaction. However, the result is that under hot alkaline washing conditions, hydrolysis occurs, leading to the stripping of dye from the fibre, and, because of liberation of the active form of the dye, there is some reaction with adjacent white cellulosic material, giving rise to a stain.

Clearly the ideal reactive dye for cellulosic materials has yet to be found, and intensive research on a world-wide scale is continuing. The economic basis for these efforts is the quite remarkable success achieved by this new class of dyes due to their very wide gamut of shades, simplicity of application, and not least their suitability for numerous advanced techniques of application, leading to quite startling increases in productivity.

REFERENCES

[1] White, G. S. J. *J. Soc. Dy. Col.*, **76**, 16, 1960.

[2] Gund, F. 'Symposium on Textile Printing', p. 147. Society of Dyers and Colourists. 1953.

[3] Cross, C. F. and Bevan, E. J. 'Researches on Cellulose', p. 34. 1895.

[4] Briggs, G. J. *Z. angew. Chem.*, **26**, 256, 1913.

[5] Schroeter, G. *Ber. dtsch. chem. Ges.*, **39**, 1570, 1906.

[6] Dirklage, R. British Patent Specification 389279.

[7] Riesz, E. *Bull. Soc. industr. Mulhouse*, **99**, 349, 1933.

[8] Ciba. British Patent Specification 363897.

[9] Günther, F. U.S. Patent Specification 1567731.

[10] Alsberg, F., Rattee, I. D., Stephen, W. E., Marshall, W. J., Speke, R. W., and Weston, C. D. British Patent Specification 797946.

Alsberg, F., Heslop, R. N., Rattee, I. D., Stephen, W. E., Marshall, W. J., Speke, R. W., and Weston, C. D. British Patent Specification 798121.

[11] Farbwerke Hoechst. British Patent Specification 733471.

[12] Peacock, D. H. *J. Soc. Dy. Col.*, **42**, 53, 1926.

[13] Granacher, C. British Patent Specifications 346385, 347117, 347263.

[14] Dreyfus, H. British Patent Specification 344420.

[15] I.G. Farben. German Patent Specification 492062.

[16] Pancirolli, F. *Boll. Rep. Fibre tess. Veg.*, **32**, 314, 1937.

[17] Guthrie, J. D. *Amer. Dyest. Rep.*, **41**, 13, 30, 1952.

[18] Ciba. British Patent Specification 587467.

[19] Dawson, T. L., Preston, C., and Fern, A. S. *J. Soc. Dy. Col.*, **76**, 210, 1960.

[20] Sause, R. (Unpublished work.)

[21] Preston, C. and Fern, A. S. *Chimia*, **15**, 177, 1961.

[22] Sumner, H. *J. Soc. Dy. Col.* (In preparation.)

[23] *Idem. Ibid.*, **76**, 672, 1960.

[24] Neale, S. M. *J. Text. Inst.* **20**, T373, 1929.

[25] Danckwerts, P. V. *Trans. Faraday Soc.*, **46**, 300, 1950.

[26] Gardner, T. S. and Purves, C. B. *J. Amer. chem. Soc.*, **64**, 1539, 1942.

[27] Vickerstaff, T. *Amer. Dyest. Rep.*, **47**, 33, 1958.

[28] Wegmann, J. *Melliand Textilber.*, **39**, 1006, 1958.

[29] Schwertassek, K. *Faserforsch. u. Textiltech.*, **9**, 321, 1958.

Greek fire

J. R. PARTINGTON

The fall of Constantinople in 1453 marked the end of the Middle Ages and was an event of profound historical significance. Before this, Greek fire had proved of great value to the Byzantine empire in its centuries-long resistance to a host of enemies. Despite the military importance of this incendiary substance, which seems to have been due as much to its terrifying appearance as to its destructive powers, its chemical nature has long been controversial. In this article convincing evidence is advanced that Greek fire was neither gunpowder nor a mixture of quick-lime and sulphur, but a light petroleum fraction thickened by addition of resin or sulphur. The traditional attribution of the discovery to Kallinikos is not conclusive.

The use of some form of fire in combat goes back to very early times. The Assyrians used burning pitch or crude petroleum in defending the walls of cities in siege warfare. Fireships were used by the Greeks from the fifth century B.C., but incendiary arrows, still used by undeveloped races, came into common use in Greek armies only after the death of Alexander the Great. Appollodorus of Damascus (A.D. 60–130), Hadrian's architect, describes an elaborate apparatus that used bellows for throwing burning charcoal powder on to stone walls, which were then cracked by splashing them with vinegar. Roman troops used incendiary arrows tipped with a mixture of sulphur, resin, bitumen, and petroleum; large spears with iron heads tipped with this incendiary mixture were thrown in sieges by engines mounted on wooden towers.

Petroleum was certainly well known in antiquity, and was obtained from Parthia, Media, Persia, and the region around the Caspian and Black Seas. Many of the sources now yielding petroleum in these regions were known and worked in quite early times; some which are mentioned in ancient accounts have been rediscovered fairly recently. The name *naphtha* used by the Greeks, is derived from the old Persian name *naft*. Strabo reports that when Alexander the Great was in the city of Ecbatana in Media he was shown a demonstration in which naphtha was sprinkled on a street and ignited at one end, when the flame flashed in an instant to the other end. Pliny tells us that liquid bitumen from Babylon was burnt in lamps. There was a 'white naphtha' which was perhaps the product of filtering cruder material through adsorbent earth or, soon after the beginning of the Christian era, of distillation. Oil of turpentine and mercury were prepared by a crude form of distillation in which the vapour from a material heated in a pot was condensed on a cooled lid; but true distillation, in which the still is separated from the receiver by a tube, was an invention attributed to Mary the Jewess, an Egyptian alchemist of the early Christian era. From Egypt, distillation apparatus passed to Syria and the Arabs, and to Constantinople, where chemistry was taught by Stephanos in the first half of the seventh century. Glass still-heads of the fifth to eighth centuries, excavated in Syria, are of the same form as those shown in Greek alchemical manuscripts and attributed to Mary. Crude petroleum, probably from the Caspian region, was imported into Constantinople and used on a very large scale for heating baths. Although we have no definite information about the distillation of petroleum it seems very likely, in view of the history of the process and the availability of the raw material, that light petroleum was obtained by distillation before the eighth century.

A different kind of incendiary material was that called 'automatic fire'. It was composed of quicklime mixed with easily inflammable materials such as sulphur, resins, powdered pyrites, and petroleum. In contact with water, the heat evolved by the hydration of the quicklime, it was supposed, would be sufficient to ignite the mixture, which was to be put on or under the wooden roofs of buildings, or on wooden catapults, there to be ignited by dew or rain. A recipe for this automatic fire is given in a sixth-century interpolation of a work, *Kestoi*, written in Greek by Julius Africanus, who was born about A.D. 170 in Jerusalem and was an engineer in the service of the Emperor Alexander Severus. The recipe says the mixture was stored in bronze boxes with airtight lids, and for use was secretly smeared over the engines of the enemy in the evening: 'when the sun rises all will be burnt'. The idea of this automatic fire seems to have come from a story in Livy, saying that in the Bacchanalia of 186 B.C. torches with heads composed of quicklime and sulphur were dipped into water and withdrawn in flames. A Chinese account of a river battle of A.D. 1161 says that 'thunderbolt missiles' consisting of paper bags filled with quicklime and sulphur were thrown on the water, when they burst into flames and leaped upwards. Nevertheless, it seems improbable that such compositions would act effectively or were ever actually used in warfare. The mixture of sulphur and quicklime has been identified with Greek fire, but the properties and mode of use of the latter were quite different.

THE BYZANTINE AND ARABIC CONTRIBUTIONS

True Greek fire was invented in Constantinople, for long centuries a bulwark against Asiatic conquest; the Byzantines called it 'liquid fire', the misleading name *feu grégeois* being coined by Crusaders. The Byzantines did not call themselves Greeks but regarded themselves as Romans, the name Greek at that time being rather a term of reproach. Both Byzantine and Arab chroniclers agree that from the end of the seventh century an incendiary far more terrible than anything known before came into use in sea battles. Byzantine Constantinople was long able by diplomacy and arms to resist attacks by Arabs, Persians, Goths, Huns, Normans, Russians, and Franks; but it fell at last, in 1453, to the Turks, whom it had resisted for centuries.

The Byzantine culture was refined and active—it preserved the heritage of the older civilizations of the West—and the army was well disciplined. The Roman arsenals even in their best days could have supplied no scientific or mechanical contrivances unknown to the Byzantine engineers, whose education, discipline, and practice had been perpetuated in unbroken succession from the times of Trajan and Constantine. The city was full of inventors and craftsmen, and, as we have noted earlier, the alchemist Stephanos lived there a little before Greek fire was invented. There was keen interest in chemistry, and in the seventh and eighth centuries Greek chemical treatises were collected and edited. In the ninth century the philosopher Leo of Thessalonika made for the Emperor Theophilus a golden tree, whose branches carried artificial birds that flapped their wings and sang; a model lion that moved and roared; and a bejewelled clockwork lady who walked. While of no importance in themselves, such contrivances are evidence of the evolution of a considerable degree of technical skill. In such circumstances the invention of Greek fire and of effective means of using it in warfare need occasion no surprise.

Theophanes in his *Chronography*, written in 811–15, says that in 671 the Emperor learned that the Arabs were preparing to attack the city and had established their winter quarters in Asia Minor. In reply to the threat he ordered the assembly of a fleet of fireships equipped with 'siphons', fire-projecting devices to whose precise nature we will turn later. In the spring the siege commenced and was carried out during the summer months for seven years. Finally, the Arab boats and sailors were destroyed at Kyzikos, partly by a storm and partly by a form of fire alleged to have been invented by the architect Kallinikos of Heliopolis in Syria, who had fled to the Romans.

The accounts of the sieges given by Byzantine and Arab historians are, as Gibbon and others have pointed out, confused. It seems, however, that there were five Arab expeditions in all—three in 665, 668, and 674; one in 715, when Greek fire was again used; and one in 782. The Arabs only once made a landing, but failed to capture the city; in 782 they again ventured within sight of Constantinople, but again turned away.

The contribution of Kallinikos to these events is uncertain, and the association of his name with the invention of Greek fire may be unjustified. Theophanes says that in 671 fireships were fitted with siphons, and it may be true that Kallinikos provided a new kind of incendiary mixture that made these more effective. Nevertheless, it is clear that at that time Byzantium already contained knowledgeable Greek chemists perfectly capable of making the invention for which Kallinikos received credit in official circles. Nearly two centuries later, the Emperor Constantine Porphyrogenitus (905–59) said that the recipe for the fire had been revealed by an angel to Constantine the Great, and that those making it known were anathema and would die instantly in disclosing it. Porphyrogenitus, who lent troops and military engines to his allies, reserved the secret of its manufacture and sent the fire ready-made to his faithful but not wholly reliable dependants. But Greek fire was not a prerogative of either angels or emperors; it had been discovered by chemists, and further east, in Saracen lands, other chemists were soon busy with its independent discovery.

The way in which Greek fire was used in naval battles is described in a treatise on tactics and strategy attributed to the Emperor Leo, either the Isaurian (A.D. 717–41) or the Armenian (A.D. 886–911). He says: 'Many valuable military engines of destruction have been invented by the ancients and moderns for attacking ships and sailors. Of this kind is the emission of fire with thunder and burning smoke, which burns ships'. The context indicates that 'thunder' does not denote an explosion; the ships were burnt and not blown up. The story has grown in the telling, and later exaggerated descriptions, which reach a climax in Oriental accounts, have misled many into thinking that gunpowder must have been meant.

Leo goes on to speak of 'the other device of the small siphons discharged by hand from behind iron shields, which are called hand-siphons and have recently been manufactured in our dominions . . . these can throw the prepared fire into the faces of the enemy'. The allusion was thought to have been to grenades filled with gunpowder thrown by troops, but their nature is not now in doubt. The term siphon was commonly used for a pump, and a small hand-pump projecting Greek fire is shown in a tenth-century Byzantine manuscript. In the next century Abu'-l-Qāsim, after describing and depicting in his book on surgery a cylindrical syringe with a piston, says it is 'like the tube by means of which naphtha is thrown in sea-combats'.

Leo further describes how the Greek fire was projected from a ship: 'The front part of the ship had a bronze tube so arranged that the prepared fire could be projected forward, to left or right, and also made to fall from above. This tube was mounted on a false floor above the deck on which the specialist troops were accommodated, and so raised above the attacking forces assembled in the prow. The fire was thrown either on the enemy's ships or in the faces of the attacking troops'. In this case also we have a picture, although not a very accurate one, of the contrivance in a tenth-century manuscript (figure 1). There seems to be no doubt that the essence of it was that a flaming liquid was projected by a force-pump through a metal tube.

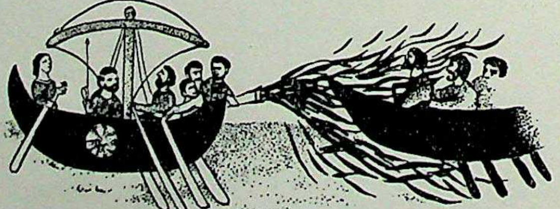


FIGURE 1 - *Use of Greek fire in a naval battle. (From a tenth-century Byzantine manuscript.)*

Anna Komnena, daughter of the Byzantine Emperor Alexios I Komnenos, describes a naval battle between the Greeks and Pisans near the island of Rhodes in 1103. Each of the Byzantine galleys was fitted in the prow with a tube ending with the head of a lion or other beast made of gilded brass or iron, frightful to behold, through the open mouth of which fire could be projected by means of a flexible apparatus. The technique was to ram an enemy ship in the stern and to pump the fire over it. The Pisans fled, having no previous experience of this device, wondering that fire, which usually burns upwards, could be so directed downwards or towards either side, according to the will of the engineer who discharged it. This account agrees exactly with that given earlier by the Emperor Leo. Apparently the projection apparatus was made to swivel, the open mouths of the fixed beasts being wide enough to allow for this.

Anna also mentions an incendiary blowpipe, in this case apparently fed with a powder or paste projected by the breath past a kindling device on the tip of the tube; it seems to have done little damage, though the effect was formidable. She says that in an encounter in an underground sap, at the siege of Durazzo (1108), the Normans had their beards singed by the Byzantines but were not otherwise much harmed.

After their experience before Constantinople, the Arabs, to whom petroleum was quite familiar, soon made Greek fire for themselves with the help of their own very expert chemists. Under the Abbasid rulers, of Persian origin, the Muslim armies were equipped with 'naphtha troops' who wore fireproof suits and projected the burning liquid by special pumps. In the tenth century a 'fierce fiery oil' was brought to China by Arabs, and soon after a special apparatus was used by the Chinese to project it by a pump.

Richard I, in his voyage from Cyprus to Acre, captured a Saracen transport ship laden with all kinds of weapons and an abundance of Greek fire in bottles, which an eye-witness had seen put aboard at Beyrut. Pictures of such ships, stacked with spherical pots or bottles, are to be seen in a thirteenth-century Arabic manuscript. In the twelfth century, surprisingly large quantities of petroleum were available to the Arabs: 20 000 barrels of it were used to burn down the city of Cairo to prevent its recapture by the

Franks. Arabic accounts refer to the large stocks of naphtha held also by the Byzantines, who were alleged to store it in churches.

To the Crusaders, the Greek fire used by their Saracen opponents was something quite new and unexpected. The siege of Mansura in the Nile Delta in 1249, in which King Louis IX of France was captured by the Mamluk Sultan, is described by Jean de Joinville (1224–1319), in whose time gunpowder was known and cannon were imminent. The Saracens discharged the fire, burning in tubs, from catapults: 'It was like a big tun and had a tail of the length of a large spear: the noise which it made resembled thunder and it appeared like a great fiery dragon flying through the air, giving such a light that we could see in our camp as clearly as in broad day'. When the French commander saw the Saracens preparing to discharge the fire he announced in panic that they were irretrievably lost.

The Crusaders believed that anyone struck by Greek fire was doomed, not only because it could kill instantly, but because its 'fall-out' was a deadly poison, as even trifling gunshot wounds were later supposed to be. Yet among all the terror Joinville does not produce a single fatal casualty due to Greek fire. Clearly its effects were greatly exaggerated. There are some interesting Arabic treatises on incendiary warfare of the twelfth and thirteenth centuries which contain highly coloured accounts of the use of distilled naphtha. It was said that if it was thrown in jars against a castle, the exhalation of the mixture would inflame and the castle would crumble with a noise of thunder and a frightful whistling. Any debris left could be destroyed by a prepared water and naphtha, which burned with a thick black smoke, lethal by its terrible and suffocating stench. That such gross exaggeration could be current is evidence of the terror that Greek fire aroused and of the moral advantage to its possessors.

THE COMPOSITION OF GREEK FIRE

The nature of Greek fire, which has been the subject of research and argument for a long time, seems to follow almost automatically from the above descriptions of its use and effects. It was a liquid, prepared by some process, that was either projected burning by pumps, especially in sea warfare, or in land warfare was commonly thrown in hand grenades or, burning in large tubs, by catapults. The pumps throwing it were called siphons; hand siphons were small pumps carried by soldiers.

Jacques de Vitry and Gossuin of Metz, in the period of the Crusades, significantly say that Greek fire was made from a liquid obtained from springs in the East, and this supports the view to which we have already briefly alluded, that its basis was a light petroleum fraction obtained by distillation. Petrol by itself would not be very effective when used in flame-projectors, as the jet would dissipate too rapidly and would not carry far enough. To overcome this technical defect it was probably thickened almost to a jelly by dissolving in it resins and perhaps also sulphur. Such a glutinous mixture would cling to its target and so be more destructive. The particular mixture, or mixtures, used and the mechanical means of projecting it, together constitute the invention of Greek fire.

The composition of Greek fire may be further inferred from some accounts compiled long after its first use. One recipe is given in the so-called 'Book of Fires of Mark the Greek', usually called the *Liber Ignium.* Although this has long been supposed to be a Latin translation of a Greek original, as the name suggests, I have shown that it is practically certain that no Greek original existed, and that the document was compiled in Spain, between 1250 and 1300, probably by a Jew, from Arabic sources. The recipe reads: 'You will make Greek Fire in this way. Take live sulphur, tartar, sarcocolla and pitch, boiled salt [*sal coctum*], petroleum oil and common oil. Boil all these well together. Then immerse in it tow and set it on fire. Then kindle the fire, which is not extinguished except by urine, vinegar, or sand'. Although it has been argued that 'boiled salt' might be saltpetre, there is little doubt, as other sources show, that it was common salt refined by boiling with water and evaporation, and the name is in fact never used for saltpetre. The salt would give a yellow colour to the flame and make it look very hot.

Another recipe occurs in an Arabic manuscript composed for Saladin about 1193; it specifies a 'naphtha' for throwing that was composed of petroleum, tar, sulphur, and resin. A third recipe is given in an Arabic manuscript copied from an original of about 1225. According to this, petroleum is first purified and made white by treatment with alum, white of egg, and other substances, followed by filtration through sand. It is distilled in a cucurbit and alembic, and it is noteworthy that, in the process of distillation said to be used in Basra, the distilled naphtha is redistilled in a vessel by burning under it reeds 'one by one', that is, by a very gentle heat, which indicates that a low-boiling petroleum fraction was involved. A mixture of this with resin, fat, black pitch, and sulphur will burn on water.

Anna Komnena gave the following recipe. 'This fire is made by the following arts. From the pine and certain such evergreen trees inflammable resin is collected. This is rubbed with sulphur and put into tubes of reeds and is blown by men using it with a violent and continuous breath.' There is no mention of distilled petroleum, the essential ingredient, but as the daughter of the Byzantine Emperor she would have had good reasons for not revealing the whole secret.

Despite such evidence pointing to naphtha, many have identified Greek fire with gunpowder, but this does not agree with its properties, nor with the way in which it was used. Further, there is no information of any sort which could suggest that saltpetre, the essential constituent of gunpowder, was known at the time when Greek fire was being quite extensively used. Others have suggested different materials, such as the early incendiary materials and the 'automatic fire' composed of quicklime and sulphur, and have assumed that Greek fire contained all these. Being solid, such a mixture could not have been projected as Greek fire was, nor would it have had its properties. Compositions fantastic for the times of which we are speaking, such as calcium phosphide, have also been proposed. All the evidence, however, which I have set forth at length in my 'A History of Greek Fire and Gunpowder' (Cambridge, 1960), seems to indicate that saltpetre was not an ingredient, that Greek fire was quite distinct from gunpowder, and that the basic constituent was naphtha thickened by the addition of resin or sulphur.

Book reviews

SCIENCE AND TECHNOLOGY

McGraw-Hill Encyclopaedia of Science and Technology. *Vols. I–XIV, Aba–Zyt, 9040 pp.; Vol. XV, Index, 536 pp. McGraw-Hill Book Co. Inc., New York; McGraw-Hill Publishing Co. Ltd, London.* 1960. £75 11*s.* 6*d. net.*

There has long been a need for a reference book on science that would be more detailed than the large general encyclopaedias, but would yield its information more easily than a textbook. This need has been felt not only by scientists who wish to follow progress in fields other than their own, but by the general public in a society that attaches increasing importance to science. This encyclopaedia is intended to satisfy this need.

The editors had at the outset to decide on the range of subjects to be included. Their decision was that the basic matter of the natural sciences, and their major applications in engineering, agriculture, forestry, industrial biology, and food and other technologies, should be included. On the other hand, only the 'pre-professional science' of medicine, pharmacy, pharmacology, and the behavioural sciences was to be considered. Another restriction was that history and philosophy were to be included only when they were needed for the basic understanding of a scientific concept or of its technical application.

If its aims are to be fulfilled, there must be some restriction on the content of a book of this kind, and that on the range of subjects is a reasonable one. So, too, is that excluding all but the essential history and philosophy, although here the results are sometimes surprising: the entry under 'atomic bomb', for instance, gives no more than a brief description and explanation of the principles involved, with no hint of the background.

Projects of this kind create great editorial problems which are not simplified if those participating are separated by barriers of distance or language. It is, therefore, not surprising to see that all the sixty-three consulting editors are American. Less satisfactory is the fact that all but sixty of the contributors, rather more than two thousand in all, are also American. This is open to criticism on practical rather than nationalist grounds. So far as the scientific entries are concerned the point is of no great significance, for at the level of these volumes an American contributor would write much the same as any other of comparable calibre. But it does weaken the technological sections, where there tends to be too heavy an emphasis on American practice. For example, the important article on electric power generation makes no mention of practice outside the United States.

However, the choice of authors has, presumably, been governed largely by practical considerations and the need for speed. In this last respect the encyclopaedia is impressively, indeed astonishingly, up to date. The article on the Moon, for example, gives a full account of the photography of the far side, together with an illustration. The discovery of ^{129}Xe in meteorites, made only last year, is given in the article on Meteorites. The work is to be kept up to date by a series of year books, and provision has been made for constant revision of the main text.

The articles on specialized topics are extremely thorough: the texts are lucid and detailed, and the illustrations numerous and clear. Sampling shows that the articles contain, in general, the information expected, although sometimes under somewhat unexpected headings. Here the comprehensive index is very valuable, although one could have wished for an additional classified index. The articles themselves contain cross-references and short bibliographies.

The above criticisms are, however, of a relatively minor nature, such as no work of this magnitude could hope to escape. Generally speaking, the work successfully achieves its aims, and is to be welcomed. TREVOR I. WILLIAMS

THEORY OF MATRICES

Introduction to Matrices and Linear Transformations, *by Daniel T. Finkbeiner II. Pp. vii+248. W. H. Freeman & Co. Ltd, London.* 1960. 38*s. net.*

There are various methods for presenting the theory of matrices to students, and the course adopted in this book is the modern one of using the techniques of abstract algebra. The first three chapters deal with abstract systems, vector spaces, and linear transformations. The theory of matrices is then developed up to the canonical form for a linear transformation. The last two chapters deal with metric concepts (bilinear and quadratic functions, length, distance, and orthogonality) and with functions of matrices, including an application to differential equations. There are well-written short appendices on algebraic concepts and combinatorial equivalence, and the answers to a number of exercises are given.

Matrix calculus looms large in various branches of applied science, including economics, psychology, and operations research. Students of these subjects who have taken mathematics to the first-year level at the university will find the present book quite readable; for, as the author remarks, 'the major prerequisite . . . is the ability to reason abstractly, to proceed logically from hypothesis to conclusion'.

L. S. GODDARD

ASTRODYNAMICS

An Introduction to Astrodynamics, *by R. M. L. Baker, Jnr., and M. W. Makemson. Pp. xiv+358. Academic Press Inc., New York; Academic Books Ltd, London.* 1960. $7.50.

This book is really an introduction to the dynamics of the solar system. As such it is welcome, particularly as it is well suited to the needs of those who wish to obtain a detailed understanding of the motion of planetary satellites and space probes but who are not primed with astronomical knowledge. The treatment is by turns mathematical and descriptive. The descriptive sections are quite readable; the mathematical sections assume the common mathematical equipment of the science graduate, without any previous knowledge of dynamics.

The first five chapters are introductory in character. They contain a discussion of the elements of orbital theory, an historical description of the astronomy of the solar system, and conclude with an excellent section on the deduction of the principal units of distance in the solar system, and of the masses of the Moon and planets.

The remaining chapters deal with orbit determination from observations, including the improvement of orbital elements by the method of differential corrections, and an account is given of perturbation techniques. One of these chapters deals with the orbital perturbations of a near-Earth satellite. There is a section on motion in a many-body field and a brief mention of the application of perturbation methods in this case; it is a pity that a description of the solutions which have been obtained in the important case of motion in the Earth-Moon system has been omitted. The closing chapters are briefly concerned with observational techniques—both at optical and radio wavelengths—and with methods of interplanetary guidance and control.

There is an extensive glossary which will be valuable to the reader without a background in astronomy. The non-astronomical definitions should be approached with caution: some, such as 'atomic clock', are not quite precise; while others, such as 'Boltzmann's constant', are incorrect. There is a good bibliography and a total of 70 exercises.

A. P. WILLMORE

GENERAL PHYSICS

Fisica Generale e Sperimentale (*seventh edition, completely revised*), *by E. Perucca. Vol. I, pp. xxviii+920; vol. II, part I, pp. xii+618; part II, pp. xi+619–1228. Unione Tipografico-Editrice, Turin.* 1960. *Vol. I, lire* 12 000; *vol. II* (*both parts*), *lire* 16 500.

For over 25 years this has been the standard Italian treatise of experimental physics, and its ever-youthful author must be warmly congratulated in having produced this new, enlarged, edition, and on having made it completely up to date. It is a vast canvas in the old tradition, outside the present trend of handbooks written by a fairly large number of specialists. While it inevitably lacks the authority of the latter, there are many advantages in a uniform exposition from a single point of view. Moreover the treatment is extremely clear and shows vividly the immense value of a humanistic background in the approach to physics. It is impossible to write a satisfactory short review of a book of this kind. Almost at random, one may remark that it makes a good case for a consistent use of the rationalized MKS system of units; that a judicious balance has been kept in the experimental illustrations of fundamental principles between those of historical interest and their modern counterparts; and that the amount of theory is adequate, given the aim of the book. This aim is to prepare Italian university students for their examination in *Fisica Sperimentale*; it covers broadly the syllabus of the Part I Special Examination in Physics of the University of London. The emphasis, as one would expect, is not always on the same points in the two countries.

The volumes are lavishly produced, the figures being particularly plentiful and useful. L. PINCHERLE

THEORY OF VIBRATIONS

General Dynamics of Vibrations, *by Y. Rocard. Pp. xii+522. Crosby Lockwood & Sons Ltd, London.* 1960. 80*s. net.*

Until recently almost the only satisfactory treatise on the theory of vibration was that of Professor Den Hartog, of the Massachusetts Institute of Technology. The present book, which is a translation from the third French edition, will undoubtedly become a standard authority on this subject. It covers a very wide field of engineering applications, and discusses them with the aid of adequate mathematical analysis. There are applications to machinery, to the acoustics of the electronics of radio engineering, to gyroscopic theory, to electronic filters, and to the relaxation oscillations of non-linear systems. Even the shimmy of automobile wheels is included. It is regrettable that nearly 17 years have elapsed between the French and English editions, and we must be grateful to W. T. Stern, to whom the first translation of this book is due. G. TEMPLE

ELECTRODYNAMICS

Electrodynamics of Continuous Media, *by L. D. Landau and E. M. Lifshitz (translated from the Russian by J. B. Sykes and J. S. Bell). Pp. x+417. Pergamon Press, London.* 1960. 84*s. net.*

This volume forms part 8 of a course of theoretical physics, and it is of the high standard that one expects from these two distinguished physicists. It is, indeed, set forth very much in the form of a neat series of lecture notes, and what makes it particularly valuable are the excellent collections of interesting examples, for which full solutions are given.

As it is a book on theoretical physics, experimental details are, of course, omitted. The sections on electrostatics and dielectrics are particularly good, and although some sections—for example, that on the passage of fast particles through matter—are somewhat condensed, physicists will find the volume a most valuable reference book, provided they have an adequate knowledge of tensor analysis. L. F. BATES

CRYSTAL PHYSICS

Imperfections in Crystals, *by H. G. van Bueren. Pp. xviii+676. North-Holland Publishing Co., Amsterdam.* 1960. £5 10*s. net.*

The role that imperfections play in the physical properties of crystals is very important. For some, such as the mechanical strength of metals, it is decisive. The last three decades have seen a systematic study of imperfections leading to the development of a vast field of physics of immense importance to technology. The purpose of this book is to summarize as concisely as possible the theories and the facts.

The first part deals with the properties of lattice imperfections in a general manner, particularly point defects and dislocations. The second part deals in more detail with imperfections in metals, and their relevance to plastic deformation, work-hardening, creep, and so on, and their effect on such important properties as electrical and thermal conductivities. The third part deals with imperfections in non-metallic crystals, particularly the alkali halides, the silver halides, and the industrially important diamond-structure materials, silicon and germanium.

The field is already so vast that much has to be left out if the book is to be made readable. The selection has been skilfully done. The results of theoretical calculations are presented with just enough of the principles of the derivation as is necessary to follow the argument. The diagrams, however, are occasionally so condensed as to be difficult to understand. The experimental evidence is presented succinctly and well illustrated with excellent photographic plates; references to original work are plentiful; and the English is lucid. Graduate students and research workers in solid-state physics will find the book useful, and interested non-specialists will find it readable.

M. H. L. PRYCE

INORGANIC CHEMISTRY

Chemical Periodicity, *by R. T. Sanderson. Pp. x+330. Reinhold Publishing Corporation, New York; Chapman and Hall Ltd, London.* 1960. 94*s. net.*

The rapidly increasing amount of factual information accumulating in inorganic chemistry causes one to look with special interest at any new attempt to classify and correlate these data. In this book the author attempts to make special use of the concept of electronegativity for this purpose. He has published several papers on this subject and has previously drawn attention to many interesting relationships in the Periodic Table.

The book begins with a short discussion of the structure of the atom and of energy levels therein, and then deals briefly with the nature of chemical bonds. This section is followed by a review of electronegativity with a natural emphasis upon the author's own empirical electronegativity values, called stability ratios. He defines the latter as the ratio of the average electron density of the atom to that of a hypothetical iso-electronic inert-gas atom.

In Chapters III and IV topics such as bond polarity, charge distribution in bonds, bond energy, molecular structure, and some aspects of co-ordination chemistry are dealt with.

Then follows a useful chapter on the physical nature of the elements. The remainder of the book is devoted mainly to a discussion of the binary compounds of the elements with oxygen, nitrogen, the halogens, and carbon. The emphasis on these naturally limits the scope of the treatment of other types of compounds.

The chief deficiency of the book is the failure to utilize adequately other complementary principles of physical chemistry, such as the role of lattice energies, hydration energies, and the crystal (ligand) field theory. Also, the relative attention paid to certain topics is puzzling: thus the use of quadrupole coupling constants receives a full page, whereas the important ligand field theory is dismissed in a quarter of a page. There is a tendency to regard complex compounds as a topic rather than as an approach to inorganic chemistry as a whole. Several out of date explanations are encountered; thus, the explanation of square cupric complexes on page 61 is no longer acceptable, and the statement concerning Cr^{+++} complexes on the same page is not correct.

Teachers will find the tabulated data and the graphs very useful, and they will find many of the ideas stimulating and provocative. However, it is another matter whether the author has succeeded in demonstrating for students the fundamental unity of the subject to which he refers in the preface.

R. S. NYHOLM

ORGANOMETALLIC COMPOUNDS

Organo-Metallic Compounds (*second edition*), *by G. E. Coates. Pp. xiii+366. Methuen & Co. Ltd, London; John Wiley & Sons Inc., New York.* 1960. 45*s. net.*

The publication in 1956 of the first edition of this book was warmly welcomed, and since then the volume has been consulted extensively by both undergraduates and research workers. Even in four years, however, so many important developments have taken place that a revision of the text was desirable. The amount of new material has posed a problem. Although the author has been more selective than previously, this second edition is about twice the size of the first, and has outgrown the scope of the series in which it originally appeared. Most of the text has been rewritten. The largest additions concern the sections on the alkali metals, boron, tin, and, in particular, the transition elements. Valency problems, for example those concerning metal-hydrocarbon π complexes, that were mentioned only briefly in the first edition, are discussed at much greater length. In a book of this size the treatment of an expanding subject cannot be completely comprehensive, but the author has succeeded in including an enormous wealth of information. Major developments up to the autumn of 1959 are discussed, and a few references to literature appearing in 1960 have been noted. The author has done a very good job, and this new edition of his book deserves every success.

W. G. OVEREND

ORGANIC CHEMISTRY

Preparativní reakce v organické chemii. V. Aldolisace a příbuzné reakce. (Reactions in Preparative Organic Chemistry. *Vol.* v, Aldol and Similar Reactions), *by J. Plešek and A. Zobáčová.*

Pp. 976. *Czechoslovak Academy of Sciences, Prague.* 1960. *Kčs* 89.

An aldol reaction in the widest sense is any reaction in which a mobile α-hydrogen reacts with a carbonyl or similar group. There are tens of thousands of such reactions, of which 2700 are described in this massive volume, which comprises three main sections. The first is entitled Aldol Reactions and Condensations and deals, *inter alia*, with actual aldol reactions, such as self-additions, intra-molecular condensations, and aldehyde-ketone reactions. The second section is devoted to the Claisen and related types of reactions. The last section deals with Michaels addition reactions and is particularly useful for its chapter on cyanoethylation. All these syntheses are catalysed by bases; but acid-catalysed reactions, such as the Mannich ones, are included at the end of the first section.

2664 references are cited, including patents, and cover the world literature, mainly from 1950 to 1957; there is no bias in favour of the literature of any country. An author and subject index of 35 pages adds to the usefulness of the book. This type of book, where every page is covered with organic formulae, is notoriously difficult to produce, and one must congratulate the editors for having done an excellent job.

M. K. SCHWITZER

STRUCTURAL ORGANIC CHEMISTRY

Optical Rotatory Dispersion, *by C. Djerassi. Pp. xiii* + 293. *McGraw-Hill Book Co. Inc., New York; McGraw-Hill Publishing Co. Ltd, London.* 1960. 74*s. net.*

The phenomenon of optical rotatory dispersion—the change of optical rotation with wavelength—has been known for nearly 150 years. The use of this property in structural organic chemistry is, however, very recent. Its application, which has been due to the introduction of new instruments, has been almost entirely the work of one school, that of Professor Carl Djerassi, of Stanford University, California.

This book is the first one on the subject that has been written from the point of view of the organic chemist. It presents a very thorough and yet at the same time a most readable account of the subject, told in a pleasantly personal style. Much of the book is rightly devoted to Djerassi's own work, which has been largely in the field of steroids, terpenes, and related alicyclic compounds. Chapters on other subjects have been prepared by experts in the fields. These chapters are as follows: instrumentation (A. N. James and B. Sjöberg), theory (A. Moscowitz), polypeptides and proteins (E. R. Blout), and amino acids (J. A. Schellman).

Many natural-product chemists will need no introduction to this book. Optical rotatory dispersion is not, however, a specialized tool for some types of natural products; it can be of service in any field of chemistry, organic or inorganic, in which asymmetric compounds occur. Rapid progress is being made at present both in the technique and in its applications.

The production of the book is excellent, and the author, publishers, and printers all deserve congratulation on the speed with which the book was produced.

W. KLYNE

GEOCHEMISTRY

Methods in Geochemistry, *edited by A. A. Smales and L. R. Wager. Pp. vii* + 464. *Interscience Publishers Inc., New York; Interscience Publishers Ltd, London.* 1960. $13.50.

This book is an extremely useful guide to the theory, techniques, and applications of the analytical methods used in the collection of geochemical and geochronological data. However, the breadth of treatment implied by the title is not fulfilled, as this is a book on analytical geochemistry only, and does not include other methods, such as those of experimental geochemistry.

The analytical techniques described include refinements of the classical wet chemical methods, colorimetric and flame-photometric methods used in rapid silicate analysis, optical-emission spectrography, fluorescent X-ray spectrography, mass spectrometry (and isotope dilution analysis), radiochemical methods (with a separate chapter on radioactivation analysis), and polarography. An important early chapter describes the techniques of collection and preparation of materials for analysis, and the final chapter is devoted to chemical methods of separation by ion exchange, solvent extraction, and paper chromatography.

The contributions (by twelve authors) are, on the whole, clear and concise. The book is not intended to be comprehensive; for relevant details the reader is constantly referred to the good selection of up to date books, reviews, and papers, given at the end of each chapter.

The book should prove valuable to those who want to know how the analytical data of geochemistry are obtained and the relative values of the methods used in obtaining them.

M. STONE

CELL GROWTH MECHANISMS

Developing Cell Systems and Their Control, *edited by Dorothea Rudnick. Pp. vi* + 240. *The Ronald Press Co., New York.* 1960. $8.

The theme of the nine papers presented here is biological differentiation and the control mechanisms that may be involved. Tardent and Bonner deal with whole organisms and discuss regenerating hydroids and developing slime-moulds. Moscona describes some of his own work in one of the more promising fields of modern embryology, the reaggregation of dissociated cells. Hormonal control mechanisms in plants and animals are the subject of two papers by Phinney, West, and Villee, while Singer discusses the curious dependence of amphibian-limb regeneration on the growth of nerves into the limb bud. The last three papers go down to the fundamental mechanisms in cell and molecular biology, whose connections with differentiation are both undoubted and unknown. Work reviews protein synthesis, and Stern describes some interesting experiments on the biochemistry of the cell cycle. Novikoff's paper is a lively and stimulating account of his cytochemical work and is an example of the good that can be done in a field that is often pedestrian. But it has little direct connection with the main theme of the book, and this emphasizes the only major criticism that can be made about this symposium. Differentiation is a vast, amorphous, and very difficult subject, and it would have been better, within a limited framework of nine papers, to concentrate on one level of organization, rather than to try to bridge the gap between animals and molecules.

J. M. MITCHISON

ENZYMOLOGY

Advances in Enzymology, Vol. XXII, *edited by F. F. Nord. Pp. v* + 567. *Interscience Publishers Inc., New York; Interscience Publishers Ltd, London.* 1960. £5 6*s. net.*

All enzymologists and most other biochemists look forward to the annual appearance of 'Advances in Enzymology'.

The latest volume presents a rich diet that should disappoint nobody.

The contents of volume XXII cover a wide range, from pure enzymology (the active site and enzyme action, synthesis and hydrolysis of sulphate esters, coenzyme binding, and enzyme chromatography) to chemical metabolism and genetics (the biosynthesis of cholesterol, the biochemistry of sulphonium compounds, the induced synthesis of proteins, and genetically controlled differences in enzyme activity). We also find a most interesting article on the synthesis of nucleotide coenzymes, written by two organic chemists.

There is no question in the reviewer's mind as to the high quality of all the articles, but it does seem odd that some of the topics are included in a specialized volume on enzymology. However, the subtitle is 'and related subjects in biochemistry'. The authors write for the general biochemist and yet succeed in being helpful to the specialist. Obviously, the balance between these two aims differs somewhat from chapter to chapter.

H. GUTFREUND

COMPARATIVE ANATOMY AND TAXONOMY

Primates, Comparative Anatomy and Taxonomy, *Vol. IV*, Cebidae, *Part A*, *by W. C. Osman Hill. Pp. xxii+523. Edinburgh University Press, Edinburgh.* 1960. £9 9*s. net.*

Dr. Osman Hill's encyclopaedic work on the comparative anatomy and taxonomy of the Primates has now reached its fourth volume with the publication of Part A of his monograph on the Cebidae. This deals with four of the sub-families of the New World monkeys. Like the previous volumes, it is mainly concerned with the analysis of the external and internal anatomical characters of the various genera and species, but it also includes accounts of habitat, diet, and behaviour. One can hardly fail to be impressed with the assiduity of the author and with the immense labour that he has undertaken to sift the voluminous literature for all the information he presents in this compendious work. More than this, he displays an intimate personal knowledge of his material, and his volumes include many original new observations of importance. To have embarked on so comprehensive a series of monographs on the Primates implies considerable courage as well as sustained effort, and the results that have so far been published are to be highly commended. The text is accompanied by a profusion of illustrations of high technical quality, and the style of writing is precise and to the point. A standard work of reference of this kind, which aims at being all-inclusive, means that in some instances data are recorded on the basis of single specimens, so that it has not been possible to indicate the range of variation in the anatomical characters of each individual species; this cannot reasonably be expected. In the course of time we look forward to the completion of this encyclopaedia with successive volumes on the other subfamilies of the Cebidae, and on the Old World monkeys and anthropoid apes. Together, they will certainly comprise the most up to date work of reference on the Primates, and will be found essential not only for primatologists, but for zoologists generally.

SIR WILFRID LE GROS CLARK

HISTORY OF TECHNOLOGY

Transactions of the Newcomen Society for the Study of the History of Engineering and Technology, *Vol. XXX* (1955–56 and 1956–57). *Pp.* 281+62 *plates. The Newcomen Society, London.* 1960. 50*s. net.*

Like its predecessors, this volume of the Newcomen *Transactions* contains such a wealth and variety of papers on the history of technology that one may comment here only on those of the widest current interest.

The recent studies of Professor Derek J. de Solla Price on the Antikythera mechanism make Dr Von Bertele's paper on the 'Origin of the Differential Gear and its Connection with Equation Clocks' of special interest, for it clears up much of the confusion between a mechanical astronomical computer as part of a clock and the actual mechanical time-measuring device itself, at least in the gearing required. Is not the time now ripe for someone to give us a first-rate general history of gears and gearing?

We are especially indebted to Mr A. Stowers, whose presidential address gives us drawings and dimensions on the basis of which a more tightly reasoned analysis of the Roman artifacts at Chollerford Bridge, Chesters, Northumberland, may be made. Professor Richmond's analysis, referred to in Vol. XXIX, is less than satisfying to any engineer familiar with the actual objects. The plan of the mill shows that only a water wheel with a vertical axle can fit the data, and the known fact that the Roman garrison in this area came from Syria accounts for this early appearance of the Greek mill in Britain. One need also only set the 'hub' with its axis vertical, note the lack of any kind of bearings, and examine the carefully dressed end surfaces to recognize it as a barrel hammer, worked by a group of men with staves, intended to produce the firm masonry characteristic of the bridge abutment.

It is a pity that the long delay in appearance of these papers in print has prevented earlier and more general discussion of their excellent material.

R. S. WOODBURY

Pilkington Brothers and the Glass Industry, *by T. C. Barker. Pp.* 296. *George Allen and Unwin Ltd, London.* 1960. 40*s. net.*

An outstanding firm has found the historian it deserves. He has traced the vicissitudes through which the operation of a single 120 ft cone, built in 1826 to make crown glass, has grown into a concern that controls plate- and sheet-glass manufacture throughout Britain. Pilkingtons is still a private limited company, but family piety has not been allowed to obscure the picture of Victorian industrial relations. Thus, after an unsuccessful strike, the new wages were to be 'the lowest than can possibly be gone to', with staggered contracts for skilled workers and a 10s. fine for a day's absence without cause. Even so, the lower wages and lower standard of living on the continent were among the factors that made the Belgians their most dangerous trade rivals.

In each generation the firm has faced a new crisis of technological advance. The latest of these, the establishment of their float process for continuous production of flat glass, involved 'the chemical and physical equilibrium problems between hot liquid glass and hot liquid metal and hot atmosphere'. It took eight years, three new plants, and £4 000 000 to find the perfect solution.

T. K. DERRY

Short notices of books

(These notices are descriptive rather than critical and are designed to give a general indication of the nature and scope of the books.)

A Dictionary of Named Effects and Laws in Chemistry, Physics and Mathematics (*Second edition*), *by D. W. G. Ballentyne and L. E. Q. Walker. Pp. v + 234. Chapman and Hall Ltd, London.* 1961. 30*s. net.*

This book is concerned with effects and laws that have been named after those who discovered or studied them. The second edition contains more than 150 new definitions, and the appendix on units has been rewritten.

Physics and Archaeology, *by M. J. Aitken. Pp. x* + 181. *Interscience Publishers Inc., New York; Interscience Publishers Ltd, London.* 1961. 43*s. net.*

The author of this book has aimed at describing some of the ways in which physics has been applied to archaeology. It is intended for both the general scientific reader and the student of archeometry (measurement of archaeological materials), and the chapters are arranged so that one technique, and even one application of a technique, can be studied more or less without reference to others. The techniques described include magnetic location, resistivity surveying, and radiocarbon and magnetic dating.

Russian Journal of Physical Chemistry, *No.* 7, 1959, *edited by Ya. I. Gerasimov. Pp.* 107. *The Chemical Society, London.* £30 *per annum.*

This is the first issue of an English translation of Журнал физической Химии. The translation will be published monthly, under the editorship of R. P. Bell, and will appear approximately three months after the Russian issue. Papers in this issue include 'Thermodynamic properties of sodium compounds in the solid state', by P. G. Maslov; 'Mutual influence of substituents in *p*-disubstituted benzene derivatives', by P. P. Shorygin; and 'Kinetics of thermal dissociation of diatomic molecules', by E. V. Stupochenko and A. I. Osipov.

Small Scale Experimental Chemistry, *by T. A. H. Peacocke. Pp. xii* + 164. *Longmans, Green & Co. Ltd, London.* 1960. 8*s.* 6*d. net.*

This book aims at providing a course of experiments to teach—using the heuristic method—the principles of qualitative inorganic analysis, and a series of small-scale inorganic preparations, such as are required for the General Certificate of Education in Britain. A certain amount of traditional usage has been discarded. Thus equations are, in general, written in ionic form, and the modern versions of the formulae of such oxyanions as zincate and aluminate are used. The emphasis is on experiments on a semimicro scale, and an appendix describes methods of making some of the apparatus that is needed.

An Introduction to Transition-Metal Chemistry; Ligand-Field Theory, *by L. E. Orgel. Pp.* 180. *Methuen & Co. Ltd, London; John Wiley & Sons Inc., New York.* 1960. 25*s. net.*

The intention of the author of this book is that it will be used by students of inorganic chemistry and others seeking a general survey of ligand-field theory, and the book is therefore both short and non-mathematical. The various chapters describe, among other subjects, orbitals and energies, stereochemistry, reaction mechanisms and reaction rates, and the complexes formed by unsaturated hydrocarbons.

Advanced Organic Chemistry (*third edition*), *by G. W. Wheland. Pp. xi* + 871. *John Wiley & Sons Inc., New York; John Wiley & Sons Ltd, London.* 1960. £7 *net.*

This textbook aims at giving, in considerable detail, structural theory in its broadest sense; it deliberately omits discussion of descriptive organic chemistry. The author has continued to emphasize resonance theory in this edition, and he has added such topics as electronic and nuclear magnetic resonance, conformational analysis, and the Hammett rho-sigma relations. He has also extended the discussion of reaction mechanisms and indicated the value of kinetic data to his approach.

Name Index of Organic Reactions (*second edition*), *by J. E. Gowan and T. S. Wheeler. Pp. iii* + 293. *Longmans, Green & Co. Ltd, London.* 1960. 50*s. net.*

A large number of organic chemical reactions are referred to by the name of their discoverer, and this book has been produced with a view to combating the obscurity that this practice can produce. The second edition is considerably larger than the first, and aims at being comprehensive; it also now includes formulae. An index of reaction types is included.

The Chemistry of the Terpenes, *by A. R. Pinder. Pp. vii* + 223. *Chapman and Hall Ltd, London.* 1960. 50*s. net.*

The aim of this book is to be a bridge between general textbooks of chemistry and comprehensive books on terpenes. Particular stress has been laid on the reactions involved in degradation and synthesis, the chemical transformations that occur in terpene chemistry, and some recent developments in stereochemistry and biosynthesis. The subjects covered include the various monoterpenes, diterpenes, triterpenes, tetraterpenes, rubber, and the biogenesis of terpenes.

Aerosols: Science and Technology, *edited by H. R. Shepherd. Pp. xiv* + 548. *Interscience Publishers Inc., New York; Interscience Publishers Ltd, London.* 1961. £8 9*s. net.*

Essentially, this book is a survey of present practice in aerosol technology. There are chapters dealing with the theory and practice of aerosols, and with the containers, valves, and propellents used. Other chapters are concerned with materials that are prepared nowadays in aerosol form, such as cosmetics, pharmaceutical products, foods, pesticides, and paints.

Biographical Memoirs of Fellows of The Royal Society, *Vol. VI. Pp.* 295. *The Royal Society, London.* 1960. 30*s. net.*

These memoirs are written by various authors and aim at giving a biography of the subject, together with an assessment of his work. The current volume includes memoirs of A. C. G. Egerton by D. M. Newitt, I. M. Heilbron by A. H. Cook, Max von Laue by P. P. Ewald, and F. A. Paneth by H. J. Emeléus.

Notes on contributors

H. BONDI,
M.A., F.R.S.

Was born in Vienna in 1919 and was educated at the *Realgymnasium*, Vienna, and at Trinity College, Cambridge. In 1945 he took up his Fellowship at Trinity College and was appointed Assistant Lecturer and later Lecturer in Mathematics in the University of Cambridge. He worked with T. Gold, F. Hoyle, and R. A. Lyttleton on problems of astronomy and cosmology. Since 1954 he has been Professor of Applied Mathematics at King's College, London.

D. J. K. O'CONNELL,
S.J., D.Sc.

Was born in England and was educated in Ireland. He was at Harvard College Observatory from 1931 to 1933; he then went to Riverview College Observatory, Australia, and was director there from 1938 until 1952, when he was appointed director of the Vatican Observatory, Castel Gandolfo. His main research work has been in the field of variable stars. Since 1955 he has been president of the Commission on Eclipsing Binaries of the International Astronomical Union.

CARL DJERASSI,
Ph.D., D.Sc.

Was born in 1923 and was educated at Kenyon College and the University of Wisconsin. He has served as a research chemist (1942–43, 1945–49) with Ciba Pharmaceutical Products, Summit, New Jersey, and later as associate director of research (1949–52) and research vice-president (1957–60) of Syntex S.A. in Mexico City. He was professor of organic chemistry at Wayne State University, Detroit, from 1952 to 1959, and since 1959 he has held his current position of professor of chemistry at Stanford University. He is a member of the U.S. National Academy of Sciences and has received the American Chemical Society's Award in Pure Chemistry (1958) as well as its Baekeland Medal (1959) and Fritzsche Medal (1960), and is the recipient of honorary D.Sc. degrees from Kenyon College and from the National University of Mexico. He has published over three hundred articles dealing with the chemistry of natural products (antibiotics, alkaloids, steroids, and terpenoids) and with applications of physical measurements (notably optical rotatory dispersion) to organic chemical problems.

J. L. SIRLIN,
Dr. Nat. Sc.

Was born in Buenos Aires in 1926, where he obtained his Dr.Nat.Sc. In 1953 he came, as a British Council scholar, to the Institute of Animal Genetics, University of Edinburgh, and after working for the British Empire Cancer Campaign became in 1960 a member of the staff of the Agricultural Research Council Unit of Animal Genetics at the Institute. He has studied genetics, embryology, and animal reproduction, and now works on biochemical cytology of protein synthesis.

I. D. RATTEE,
B.Sc., A.R.C.S.

Was born in 1926. He was educated at the Royal College of Science, London, and has worked with Imperial Chemical Industries Ltd., Dyestuffs Division, Manchester, since 1945. From 1945 to 1950 he was concerned with problems associated with the dyeing of wool and published a number of papers, as well as being involved in technical service to the wool-dyeing trade. In 1951 he commenced work on reactive dyes for wool as a long-range research project, and in 1953 was associated with Dr W. E. Stephen with the developments leading to Procion dyestuffs. For this work he and Dr. Stephen were each awarded the Gold Medal of the Society of Dyers and Colourists in 1960. Since 1953 he has been concerned with reactive dyes for cellulosic fibres, both from the research and technical points of view, and has been associated with numerous patents.

J. R. PARTINGTON,
M.B.E. (Military), D.Sc.

Was born in Bolton, Lancashire, and studied in the Universities of Manchester and Berlin. He became an assistant lecturer in Manchester University, and was, from 1919, the first University Professor of Chemistry in East London College, later Queen Mary College, University of London; he retired in 1951. He served as an officer in the Royal Engineers in the First World War. He has been a Vice-President of the Faraday Society, President of the British Society for the History of Science, and is a *Membre Effectif de l'Académie Internationale d'Histoire des Sciences*. He has published papers on inorganic and physical chemistry, on the history of chemistry, and has also written some books, the latest being 'A History of Greek Fire and Gunpowder'.

An international council of sciences

Traditionally, science knows no frontiers, and a free exchange of ideas between men of science in different lands has always been looked upon as of the first importance. In the past, this freedom has transcended the bitterest national antipathies, even war itself. Completely free exchange is no longer practicable, but the principle remains. Today there is virtually no major field of science in which research is not being pursued in several different countries, and—to the advantage of all—new knowledge and new ideas are widely and quickly exchanged through the scientific journals, through conferences, or by private communication between individuals. For many purposes, such arrangements meet the needs of the day, and there are strong arguments against trying to establish a more rigid organization: in science, as in every human endeavour, much is sacrificed when the human touch is lost. There are, nevertheless, many instances of international co-operation in which a clearly defined allocation of responsibilities is not merely desirable but essential.

The ninth General Assembly of the International Council of Scientific Unions (ICSU), held in London last month, provided a timely reminder of the great and growing importance of the existence of formal machinery for international collaboration in certain kinds of scientific projects. The International Geophysical Year, organized by a special committee of ICSU, was by far the largest scheme of international co-operation in science that has ever been carried out. Sixty-seven countries took part and more than two thousand observing stations, among them fifty-seven in the Antarctic, were established. Although a scientific project of such magnitude can only occasionally be undertaken, it is a forceful indication of what can be achieved by joint international effort. No less important, it is a welcome reminder that close and friendly scientific co-operation is possible even between countries whose political differences seem irreconcilable.

The International Geophysical Year was of further significance in relation to ICSU; it was a powerful demonstration of the Council's ability to initiate a really great enterprise. It should therefore inspire the confidence necessary to secure adequate support for a further expansion of its activities. Such an expansion seems very necessary, for since the last war there has been an immense expansion in almost every branch of science, accompanied by increasing co-operation between different scientific disciplines and between scientists of different countries. It is clear that this fruitful trend will continue; moreover, the very existence of an efficient and adequately endowed organizing body will in itself do much to encourage it. ICSU does not itself collect and disburse funds for research. Its function is to co-ordinate the work of others and fit it into an effective pattern; it undoubtedly catalyses much research that would not otherwise be done.

ICSU has grown to its present stature from modest beginnings, that can be traced back to the second half of the nineteenth century, when there existed a number of associations of rather highly specialized natures. Some were organized inter-governmentally, but many were fostered by Germany; in 1900 they were brought together as the International Association of Academies. German influence was very strong, and in fact the Association was dominated by the academies of Berlin, Göttingen, Heidelberg, Leipzig, Munich, and Vienna. The first world war almost completely put a stop to its activities; afterwards, the International Research Council was founded on the joint initiative of the *Académie des Sciences*, Paris; the Royal Society of London; and the National Academy of Sciences, Washington. This Council served as a co-ordinator for various scientific unions, including chemistry, biological sciences, and physics; but originally it was no more truly international than the old Association of Academies, for it was really limited to the Allied countries. In 1931 the Council was reconstituted as the International Council of Scientific Unions, with aims and constitution far more in accordance with its title. Its defined purpose was to co-ordinate the work of the national adhering organizations and of the various international unions. Each of the last was given complete autonomy, and the Council's function approximated to that of a federal parliament, regulating what was common to all, but not interfering in the work of individual unions.

Again war disrupted the work, but soon afterwards the Council was revived once more. Today, it has two main aims. Firstly, to co-ordinate and facilitate the activities of the unions; secondly, to serve as a co-ordinating centre for the national organizations adhering to it. At present, the fourteen unions fall into two classes—general and specialized—distinguished by their voting powers and representation. The former includes astronomy, geodesy and geophysics, pure and applied chemistry, pure and applied physics, and the biological sciences. Among the specialized unions are those for radio-science, crystallography, physiological sciences, biochemistry, and the history and philosophy of science.

While ICSU was being reconstituted, a new international agency interested in science appeared, the United Nations Educational, Scientific, and Cultural Organisation (Unesco). There are fundamental differences between them; whereas Unesco is an inter-governmental organization whose adherents are its member states, ICSU is an association of national academies of science or national research councils on the one hand, and of the international scientific unions on the other. Nevertheless, the interests of the two organizations naturally overlap, and, to avoid wasteful duplication, their respective fields of interest were defined in an agreement that has been renewed annually since 1946. Unesco provides ICSU with an annual subvention and is represented at the meetings of its Executive Board and of the General Assembly. In return, ICSU gives Unesco expert advice when called upon to do so.

The present activities of ICSU are too extensive for more than a few of the most important features to be summarized here; a full account was, however, recently written for this journal by the late Sir Harold Spencer Jones, formerly Secretary General (ENDEAVOUR, 18, 88, 1959). Apart from organizing big international projects of observation and experiment, ICSU is active in co-ordinating scientific abstracting, and maintains a system of what are called permanent services.

The Council's interest in abstracts began in 1949, when Unesco held a conference in Paris on this subject. In the same year, following its General Assembly in Copenhagen, ICSU appointed a joint commission on physics abstracting, which in 1950 took over complete responsibility for this from Unesco. In 1951, the ICSU Abstracting Board was established, and now works in collaboration with 'Physics Abstracts', *Bulletin Signalétique*, *Physikalische Berichte*, and the Institute of Scientific Information, Moscow; its guiding principle is that for any particular branch of science there should be only one abstracting journal in each language. Subsequently this interest was extended to the field of chemistry, and today the Abstracting Board co-ordinates the work of the principal chemical abstracting journals in the English, French, German, and Russian languages.

The Federation of Astronomical and Geophysical Services was established in 1956. It now numbers eleven services in all, and serves the unions of astronomy, of geodesy and geophysics, and of scientific radio. The services include an international time bureau; a gravimetric bureau; an earth tide service; and a central ursigram service for the transmission of coded ionospheric, solar, geomagnetic, and other geophysical data.

In addition, ICSU has four special committees, dealing respectively with geophysics, oceanic research, Antarctic research, and space research. At present the most important task of the committee for geophysics is to ensure the fullest possible exploitation of the results of the International Geophysical Year, but already it is making plans for further comparable projects. These include a world magnetic survey—timed for 1964–66, when solar activity will be at its next minimum—and an exploration of the mantle of the Earth. Among the projects of the special committee on oceanic research is one for a detailed survey of the Indian Ocean in 1962–63; in this, at least twenty survey vessels, provided by ten nations, will take part. The Indian Ocean has been chosen for three main reasons. Firstly, it is the least known ocean of the world. Secondly, the winds and currents can be studied through a complete monsoon cycle, during which the direction of the prevailing wind is reversed. Thirdly, the fishing areas of this ocean are of great economic importance to the countries bordering upon it. In space research, the interest of the special committee, established in 1958, is in fundamental investigation, rather than in the technological problem of the design and launching of satellites and space probes.

The effective discharge of these diverse tasks demands adequate machinery and sufficient funds to maintain it. Hitherto its total income has been no more than £115 000 [$320 000]. That so much has been achieved for so little is due to the fact that most of the work is voluntary, though naturally travel and other incidental expenses must be paid; the permanent central directorate, at the Hague, is very small. Nevertheless, the time has clearly come when the organizational tasks ahead are greater than those ICSU can undertake with its present staff and resources. At a time when international expenditure on research is measured in hundreds of millions of pounds annually, it is very much to be hoped that ICSU will find no difficulty in obtaining the relatively very modest sum required to maintain and extend its services.

The organic and biological chemistry of mevalonic acid

A. F. WAGNER and K. FOLKERS

Since its identification in 1956, mevalonic acid has been recognized as a key substance in the biosynthesis of a wide range of very important natural products, including steroids, carotenoids, and terpenes. Its discovery is therefore of major scientific importance. This article, by pioneers in the field, reviews the research that led to the recognition of mevalonic acid as a growth factor, its isolation, the identification of its structure and configuration, its synthesis, and the elucidation of its role in a number of biosynthetic processes. The final section of the article makes it clear that, in addition, mevalonic acid is of fundamental biosynthetic importance in a number of fields that have so far been only briefly investigated.

DISCOVERY AND ISOLATION

Many investigators have used materials such as distillers' dried solubles for studying factors that promote the growth of animals. A. F. Novak, S. M. Hauge, and C. W. Carrick [1] reported in 1947 that such material contained an unidentified factor necessary for the growth of chicks, and in the following year a concentrate prepared from this source was designated 'vitamin B_{13}' [2]. Reports that a 'vitamin B_{13}' concentrate would promote growth in both chicks and rats stimulated studies in other laboratories.

In 1956, H. R. Skeggs, L. D. Wright, E. L. Cresson, G. D. E. MacRae, C. H. Hoffman, D. E. Wolf, and K. Folkers [3] subjected a 'vitamin B_{13}' concentrate to countercurrent distribution, and tested various fractions so obtained for their ability to promote microbial growth. Certain fractions were found to be highly active in promoting the growth, in an acetate-free medium, of *Lactobacillus acidophilus* (strain ATCC 4963). This activity was shown to be distinct from that of lipoic acid, a growth-promoting factor obtained from liver, and of all other compounds; these studies resulted in the discovery of a new acetate-replacing factor that was designated mevalonic acid. In these studies, the growth-promoting activity of 'vitamin B_{13}' concentrates in the chick was not confirmed; W. H. Ott and his colleagues [4] considered the growth tests with the concentrates to be negative, or at best equivocal in chicks. Purification of the new acetate-replacing activity was therefore continued using assays based upon the growth of *L. acidophilus* ATCC 4963 in an acetate-free medium.

To allay suspicion that the activity might be related to a known compound, or a combination of known compounds, over 150 compounds of biological significance were tested; not one showed comparable activity in promoting the growth of *L. acidophilus* in the absence of acetate. A demonstration that the factor was not a combined form of lipoic acid was particularly significant, because this compound had been shown to be present in concentrates. When microbiological assays for both lipoic acid and the 'new factor' were shown to be unchanged after hydrolysis with normal sulphuric acid, it was concluded that the material being studied contained no conjugate of lipoic acid that was acid-susceptible and unavailable to micro-organisms. The nearly quantitative recoveries of activity in various fractionation studies were also significant, since unaccountable losses in activity would have resulted if the microbial response had been due to a combination of compounds rather than to a single one. Further experiments with concentrates of the new factor demonstrated its ability to promote the growth of *Lactobacillus bifidus*, *L. bulgaricus*, and *Thermobacterium acidophilus*.

G. Tamura [5] detected a growth factor in Japanese rice-wine (sake) and also in culture media in which organisms such as *Aspergillus*, *Penicillium*, *Monilia*, *Willia*, and *Lactobacillus* had grown. The factor was designated hiochic acid because of its activity in promoting growth of the 'true hiochi bacteria', *Lactobacillus homohiochi* and *L. heterohiochi*. On the basis of the identical infrared spectra and microbiological activities of pure compounds obtained from both sources, G. Tamura and K. Folkers [6] later reported hiochic and mevalonic acids to be identical.

L. D. Wright and his colleagues [7] obtained highly purified concentrates of mevalonic acid from distillers' dried solubles. The active substance was extracted from the material with methanol, and this methanol-soluble fraction was partially purified by continuous extraction with chloroform. The chloroform-soluble portion was purified by chromatography (on Superfiltrol), and inert material in a series of selected eluates was precipitated with ethanol. When an aqueous solution of the concentrate was passed through a cation-exchange resin (Dowex-50), the active compound ran through rapidly, close behind the solvent front, and left behind many impurities. Selected fractions from this stage were then neutralized, and the factor was adsorbed on an anion-exchange resin (Dowex-1). Selective elution from the resin with 0·05N formic acid resulted in almost quantitative recovery of the activity. These two stages of ion-exchange purification achieved a five-fold concentration. After several successive countercurrent distributions of the richest fractions between the phases of an ethanol/water/chloroform mixture, the factor was obtained as a clear oil that weight-distribution and activity curves showed to be almost pure. About 150 mg of oil, obtained from 800 kg of starting material, were thus made available for the first structural studies.

STRUCTURE DETERMINATION

In aqueous solution, the newly isolated compound behaved as an acid; on potentiometric titration of an aqueous solution with alkali, the titration curve rose sharply and then drifted, in a manner characteristic of lactones. The infrared absorption spectrum of the compound also was compatible with a lactone function. In chloroform solution, absorption bands were observed that indicated the presence of hydroxyl and ester groups. In morpholine solution, the absorption band (5·78 μ) corresponding to the ester function slowly decreased in intensity during 48 hours, and a new band (6·1 μ), corresponding to a carboxylate ion, appeared and slowly increased in intensity. Such behaviour is typical of lactones, and, on the basis of the rate of ring opening, the function was assumed to be a δ- rather than a γ-lactone.

The highly purified oil was converted to the crystalline benzhydrylamide before further structural studies were undertaken. This derivative contained one C-methyl group and one hydroxyl group that could be acylated. Elemental analysis of the benzhydrylamide and its corresponding monoacetate indicated the molecular formula $C_6H_{12}O_4$ for the active compound, and $C_6H_{10}O_3$ for the corresponding lactone. Since spectral and titration data had already indicated the factor itself to be a δ-hydroxy acid, it appeared that placing a methyl group and one more hydroxyl group on the carbon chain would satisfy the elemental composition and complete the elucidation of the structure.

Neither the lactone nor its corresponding benzhydrylamide reacted with alkaline periodate or with hypoiodite, showing that the molecule does not contain an adjacent γ-hydroxyl group or

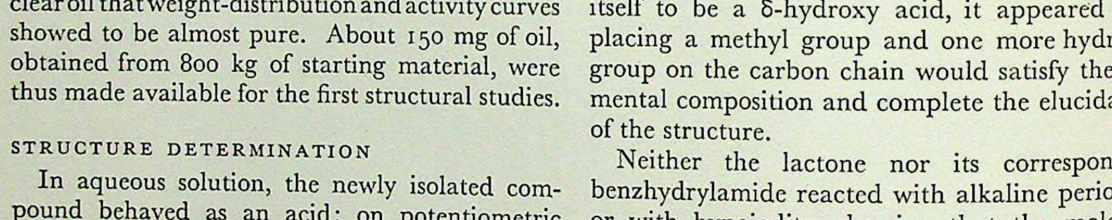

C₂H₅OOCCH₂C(CH₃)(OH)—CH₂COOC₂H₅
Diethyl β-hydroxy-β-methylglutarate (III)
LiAlH₄
HC(=O)CH₂C(CH₃)(OH)—CH₂COOC₂H₅
Ethyl γ-formyl-β-hydroxy-β-methylbutyrate (IV)
H₂
HOCH₂CH₂C(CH₃)(OH)—CH₂COOC₂H₅
Ethyl β,δ-dihydroxy-β-methylvalerate (V)
Saponify
Distil
CH₃C(=O)OCH₂CH₂C(CH₃)(OH)—CH₂COOC₂H₅
Ethyl 5-acetoxy-3-hydroxy-3-methylpentanoate (VII)
hydrolysis
HOCH₂CH₂C(CH₃)(OH)—CH₂COOH
Mevalonic acid (I)
Distil
CH₃
OH
O
O
Mevalolactone (II)
Zn (Reformatsky reaction)
CH₃C(=O)OCH₂CH₂C(=O)CH₃ + BrCH₂COOC₂H₅
4-Acetoxy-2-butanone (VI)
Ethyl bromoacetate

FIGURE 1 – *Original syntheses of mevalonic acid and mevalolactone.*

an α-hydroxyethyl substituent. Assignment of the hydroxyl group to the α-position was eliminated because of the pK of the acid; consequently, the β-position was the sole remaining possibility for the hydroxyl group.

The methyl group also was assigned to the β-position. Of the possibilities, the δ-position was eliminated as the site of the methyl group because the iodoform test was negative. Of the three remaining positions, the β-position was chosen because the behaviour of the compound on acetylation was best explained by the presence of a tertiary hydroxyl group.

On this basis, D. E. Wolf, C. H. Hoffman, P. E. Aldrich, H. R. Skeggs, L. D. Wright, and K. Folkers [8, 9] assigned the structures β, δ-dihydroxy-β-methylvaleric acid (I) and β-hydroxy-β-methylvalerolactone (II) to the acid and lactone forms of the factors respectively. The generic name mevalonic acid was subsequently used for the new biologically active factor.

SYNTHESES OF DL-MEVALONIC ACID

This structural conclusion was verified by an unequivocal synthesis [9]. The successive stages (III, IV, V, II) are shown in figure 1. The microbiological activity of the synthetic mevalolactone was one half that of the optically active substance obtained from distillers' solubles.

An improved and very useful basic synthesis by C. H. Hoffman, A. F. Wagner, A. N. Wilson, E. Walton, C. H. Shunk, D. E. Wolf, F. W. Holly, and K. Folkers [10] was based on condensation of 4-acetoxy-2-butanone (VI, figure 1) with ethyl bromoacetate. Hydrolysis of the product (VII) gave an equilibrium mixture of the corresponding dihydroxy acid and lactone. The pure lactone was isolated by distillation; the pure acid was obtained as the *bis* N,N′-dibenzylethylenediammonium salt after reacting the acid-lactone mixture with an excess of N,N′-dibenzylethylenediamine in semi-aqueous media.

SYNTHESES OF RADIOACTIVE DL-MEVALONIC ACID

Study of the role of mevalonic acid in polyisoprenoid biosynthesis required labelled mevalonic acid, and the following syntheses were utilized.

J. W. Cornforth and his colleagues [11] synthesized 2-^{14}C-mevalolactone from 4-acetoxy-2-butanone (VI) and ethyl 2-^{14}C-bromoacetate (figure 1). The synthesis of 4-^{14}C-mevalolactone [12] utilized key reactions from both of the earlier syntheses of mevalolactone. The sequence of reactions is shown in figure 2.

An excellent synthesis of 3′,4-^{13}C-mevalolactone was introduced [12] that started from doubly labelled diketene (X). The sequence of reactions is shown in figure 3.

SYNTHESIS OF (+)-MEVALONIC ACID

The biologically active form of the acid is dextrorotatory. It has been obtained by chemical resolution [13], and (−)-mevalonic acid has been obtained by enzymic resolution of the racemate [14]. (+)-Mevalonic acid has been prepared directly by enzymic synthesis [15, 16]; 70 per cent of the (+)-isomer of the racemate was isolated as phospho-(+)-mevalonic acid after selective enzymic phosphorylation with mevalonic kinase in the presence of ATP.

U. Henning, E. M. Möslein, and F. Lynen [16] accomplished the synthesis of (+)-mevalonic acid

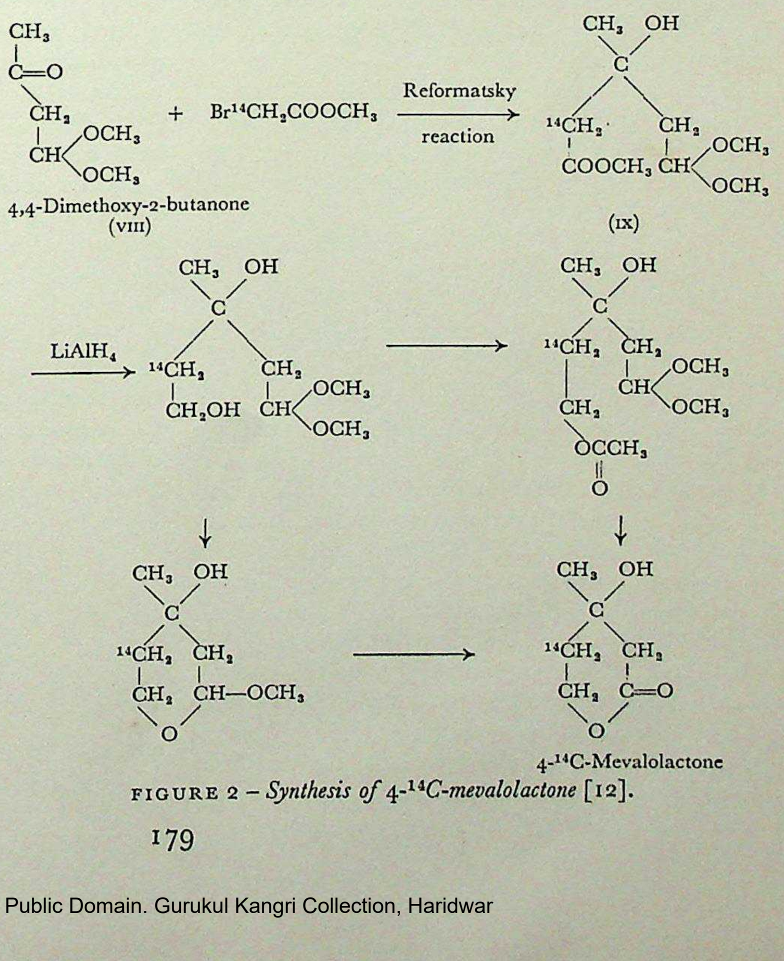


FIGURE 2 – *Synthesis of 4-^{14}C-mevalolactone* [12].

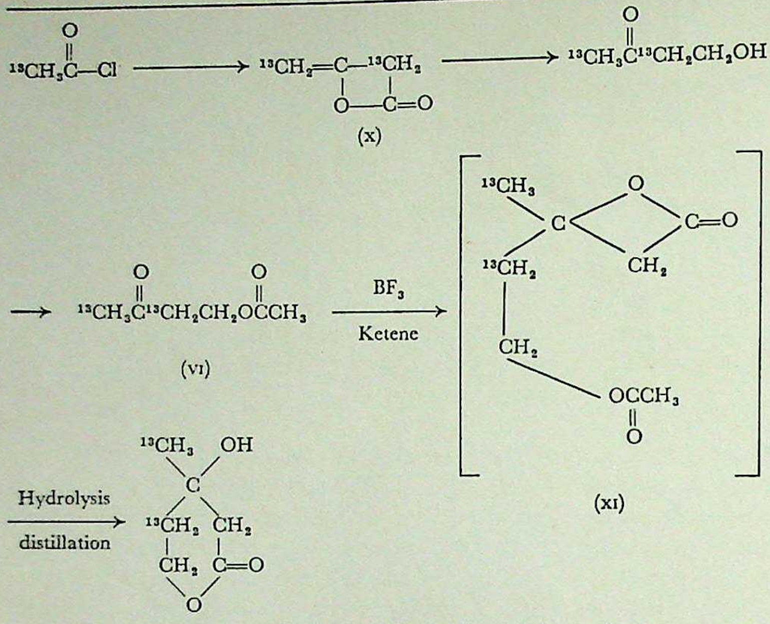


FIGURE 3 – *Synthesis of 3′,4-^{13}C-mevalolactone* [12].

from β-methylcrotonyl coenzyme A by the reaction sequence shown below, which was carried out in two stages. The carboxylation reaction was completed as a single step, but the hydration and reduction reactions were combined to prevent cleavage of β-hydroxy-β-methylglutaryl coenzyme A to acetoacetate by an aldolase present in the enzyme system.

β-methylcrotonyl CoA + ATP + $CO_2 \longrightarrow$
β-methylglutaconyl CoA + ADP + P_i
β-methylglutaconyl CoA + $H_2O \longrightarrow$
β-hydroxy-β-methylglutaryl CoA
β-hydroxy-β-methylglutaryl CoA
+ 2TPNH + $2H^+ \longrightarrow$
(+)-mevalonic acid + $2TPN^+$

CONFIGURATION OF MEVALONIC ACID

(+)-Mevalonic acid has been related to (+)-glyceraldehyde by way of quinic acid (figure 4, xii). M. Eberle and D. Arigoni [17] converted quinic acid (xii) to the 'unnatural' (+)-form of mevalolactone by the sequence of reactions outlined in figure 4. Since these reactions yielded the biologically inactive *S*-(+)-lactone (xiii), the biologically active forms of the compound are designated *R*-(−)-mevalolactone and *R*-(+)-mevalonic acid. Additional confirmation for this conclusion was obtained from a comparison of the optical-rotatory dispersion curve of the benzhydrylamide derived from the *S*-(+)-lactone (xiii) with that reported for the benzhydrylamide derived from the antipode of mevalonic acid.

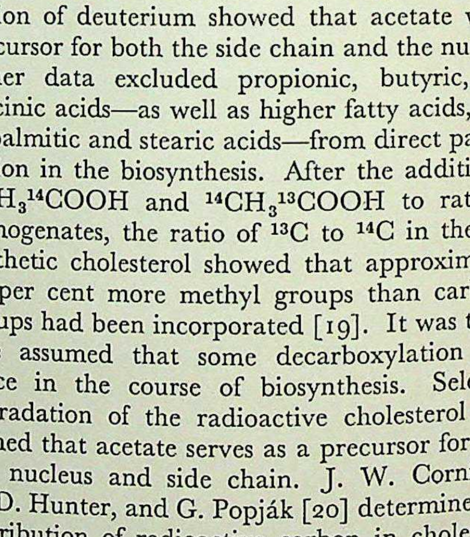
MEVALONIC ACID IN BIOSYNTHESIS

The biosynthesis of sterols

The biosynthesis of sterols and terpenes has been investigated by biochemists and biologists for the past two decades. Chronologically, significant landmarks in following the biosynthetic route (figure 5) were the discoveries that acetate, squalene, lanosterol, mevalonic acid, and isopentenyl pyrophosphate were effective precursors of cholesterol.

Acetate

The role of acetate as a precursor for cholesterol was established in animal-feeding experiments, and by using labelled acetate *in vitro*. K. Bloch and D. Rittenberg [18] observed that the ingestion of deutero-acetic acid by mice and rats led to the formation of deuterocholesterol; the distribution of deuterium showed that acetate was a precursor for both the side chain and the nucleus. Other data excluded propionic, butyric, and succinic acids—as well as higher fatty acids, such as palmitic and stearic acids—from direct participation in the biosynthesis. After the addition of $^{13}CH_3{}^{14}COOH$ and $^{14}CH_3{}^{13}COOH$ to rat-liver homogenates, the ratio of ^{13}C to ^{14}C in the biosynthetic cholesterol showed that approximately 20 per cent more methyl groups than carboxyl groups had been incorporated [19]. It was therefore assumed that some decarboxylation took place in the course of biosynthesis. Selective degradation of the radioactive cholesterol confirmed that acetate serves as a precursor for both the nucleus and side chain. J. W. Cornforth, G. D. Hunter, and G. Popják [20] determined the distribution of radioactive carbon in cholesterol biosynthesized from $^{14}CH_3COOH$ and



Cholesterol

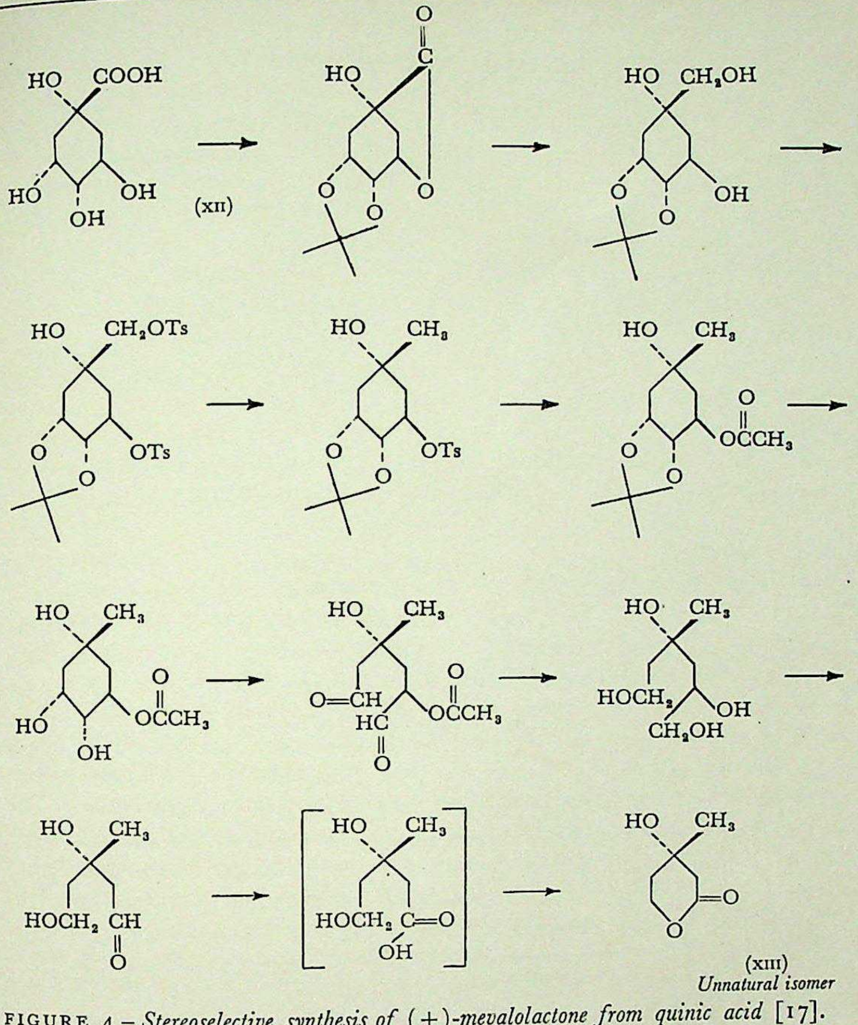


FIGURE 4 – *Stereoselective synthesis of* (+)*-mevalolactone from quinic acid* [17]. (*Ts= toluene sulphonate*).

$CH_3{}^{14}COOH$ by rat-liver slices; fifteen carbon atoms of the sterol were derived from the methyl group and twelve from the carboxyl group of acetate. The same conclusion had been reached by H. N. Little and K. Bloch [19].

Selective degradation of the biosynthetic cholesterol showed that carbon atoms 2, 4, 6, 8, and 10 were derived from the carboxyl group, and carbon atoms 1, 3, 5, and 19 were derived from the methyl group of acetate [20]. J. Wuersch, R. L. Huang, and K. Bloch [21] showed that the entire side chain of the sterol molecule was derived from acetate. After the addition of labelled acetate to an *in vitro* system and selective degradations of the side-chain, it was revealed that carbon atoms 21, 22, 24, 26, and 27 were derived from the methyl group of acetate, and that carbon atoms 20, 23, and 25 were derived from the carboxyl group. When these investigators considered the biosynthetic derivation of the carbon atoms of the side-chain on the basis of the Bonner and Arreguin mechanism [22] (figure 6) for the conversion of acetate to isoprene, it was concluded that five-carbon units related to isoprene participate in the biosynthesis of cholesterol from acetate.

The reaction scheme shown in figure 6 received further support when it was demonstrated by R. O. Brady and S. Gurin [23] that methyl-labelled acetoacetate was incorporated into cholesterol by rat-liver slices without degradation to two-carbon units. Since acetoacetate functions as a cholesterol precursor, it was suggested [24] that the next step of biosynthesis might involve condensation of acetoacetyl coenzyme A with acetyl coenzyme A to yield β-hydroxy-β-methylglutaryl coenzyme A (XIV). They demonstrated that DPN, ATP, and coenzyme A were required for the incorporation of labelled acetate, pyruvate, and acetyl coenzyme A into cholesterol by particle-free aqueous extracts of rat liver, and also observed the conversion of labelled acetate into β-hydroxy-β-methylglutarate in the same extract. H. Rudney [25] showed that radioactivity from labelled acetate was incorporated into β-hydroxy-β-methylglutaric and β,β-dimethylacrylic acids by liver homogenates and extracts. The distribution of radioactivity in the carbon skeleton of each compound was established by degradation and was compatible with the idea that such units could be utilized in biosynthesis without first being degraded to two-carbon units. A detailed investigation of the reactions leading to the conversion of acetate to hydroxymethylglutarate (XIV) conclusively demonstrated that β-hydroxy-β-methylglutaryl coenzyme A is formed from the enzymic condensation of acetyl coenzyme A and acetoacetyl coenzyme A with

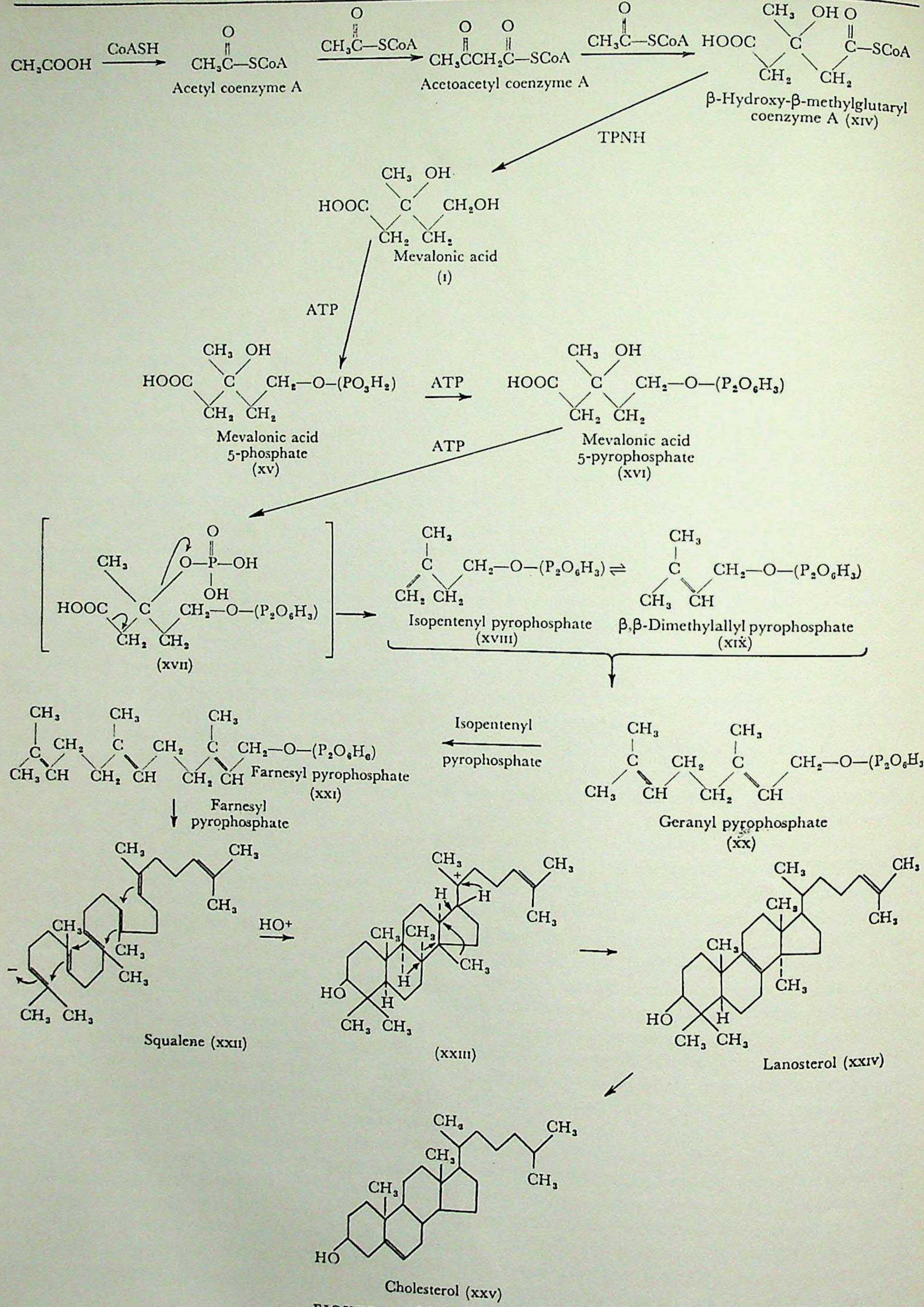

CoASH
Acetyl coenzyme A
Acetoacetyl coenzyme A
β-Hydroxy-β-methylglutaryl coenzyme A (xiv)
TPNH
Mevalonic acid (i)
ATP
Mevalonic acid 5-phosphate (xv)
Mevalonic acid 5-pyrophosphate (xvi)
(xvii)
Isopentenyl pyrophosphate (xviii)
β,β-Dimethylallyl pyrophosphate (xix)
Isopentenyl pyrophosphate
Geranyl pyrophosphate (xx)
Farnesyl pyrophosphate (xxi)
Farnesyl pyrophosphate
Squalene (xxii)
HO+
(xxiii)
Lanosterol (xxiv)
Cholesterol (xxv)

FIGURE 5 – *Biosynthesis of cholesterol.*

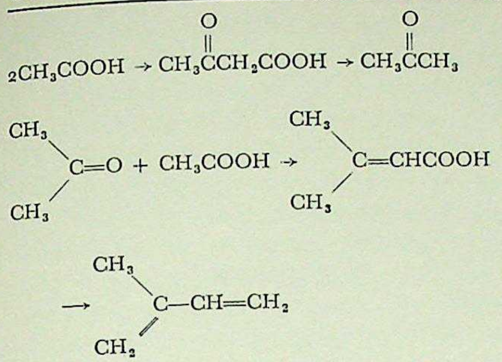


FIGURE 6 – *Mechanism for the conversion of acetate to isoprene* [22].

the concomitant hydrolysis of a thioester bond [26].

Mevalonate

Knowledge of the microbiological activity of mevalonic acid, and of the structural similarity between mevalonic acid and β-hydroxy-β-methylglutaric acid prompted experiments on the possible role of the new factor in cholesterol biosynthesis. Using cell-free rat liver homogenates, P. A. Tavormina, M. H. Gibbs, and J. W. Huff [27] showed by radioactivity experiments that 43·4 per cent of DL-2-^{14}C-mevalonic acid was incorporated into cholesterol; by contrast, only 0·16 per cent and 3·8 per cent respectively of the isotope from 3′-^{14}C-β-hydroxy-β-methylglutaric acid and 4-^{14}C-β, β-dimethylacrylic acid was incorporated into the cholesterol under comparable conditions. Assuming only one of the optical antipodes of the racemate to be utilized, it was apparent that mevalonic acid was very efficiently incorporated. It was also clear that mevalonic acid is an important biological precursor for steroids and for isoprenoid natural products in general, and that it is closely related to 'active isoprene', which is discussed later. Further studies [28] of the incorporation of ^{14}C isotope from 2-^{14}C-mevalonic acid and 1-^{14}C-mevalonic acid showed that the carboxyl group of mevalonic acid is eliminated during the biosynthesis of cholesterol; all the remaining carbon atoms are incorporated.

Mevaldic acid

Since one carboxyl group of β-hydroxy-β-methylglutaric acid is reduced to a hydroxymethyl group during the enzymic synthesis of mevalonic acid, the possible intermediate aldehydic derivative, mevaldic acid, has been studied. It was synthesized [13, 29] from the dimethyl acetal of acetoacetaldehyde and ethyl bromoacetate in a manner analogous to a synthesis of mevalonic acid (figure 1). Mevaldic acid is a very reactive aldehyde, whose presence in solution was readily established by reduction to mevalonic acid and in other ways. Although the microbiological activity of DL-mevaldic acid for *L. acidophilus* is only 0·5 per cent of that of DL-mevalonic acid, the efficiency with which it is incorporated into cholesterol by rat liver is indistinguishable from that of mevalonic acid [30].

Although mevaldic acid has not been detected as an intermediate in the enzymic reduction of β-hydroxy-β-methylglutarate to mevalonate, it has been demonstrated that it is reduced to mevalonic acid by TPNH and yeast enzymes.

Squalene

Mevalonic acid is also related to squalene (XXII), which R. G. Langdon and K. Bloch [31] showed to be a precursor of cholesterol in the intact animal; the utilization of squalene in the rat is ten to twenty times more efficient than that of acetate. J. W. Cornforth and G. Popják [32] studied the biosynthesis of squalene from labelled acetate *in vitro*, and elucidated the derivation of the carbon atoms in the squalene skeleton. The pattern of labelling in squalene was consistent with this being a direct precursor, because proper folding of labelled squalene could yield a sterol nucleus labelled in just the same way as has been established experimentally. These experiments substantially confirmed a biosynthetic pathway from acetate to sterol through squalene.

The first knowledge about intermediates between mevalonic acid and squalene was gained from studies with yeast extracts. From these investigations, mevalonic acid 5-phosphate (XV), mevalonic acid 5-pyrophosphate (XVI), 3-methyl-3-butenyl 1-pyrophosphate (XVIII, isopentenyl pyrophosphate), geranyl pyrophosphate (XX), and farnesyl pyrophosphate (XXI) have been identified as successive intermediates in the biosynthesis of squalene (figure 5).

Because ATP can be shown to be necessary for the enzymic reaction, it was suggested [33] that phosphorylated intermediates are formed in the conversion of mevalonic acid to squalene. T. T. Tchen [34] showed that the enzyme in yeast extract that converts mevalonic acid to squalene is a kinase; in the presence of ATP, this converts mevalonic acid to mevalonic acid 5-phosphate (XV). The latter was converted to squalene in the enzyme system, and it was concluded that mevalonic acid 5-phosphate is the first intermediate in

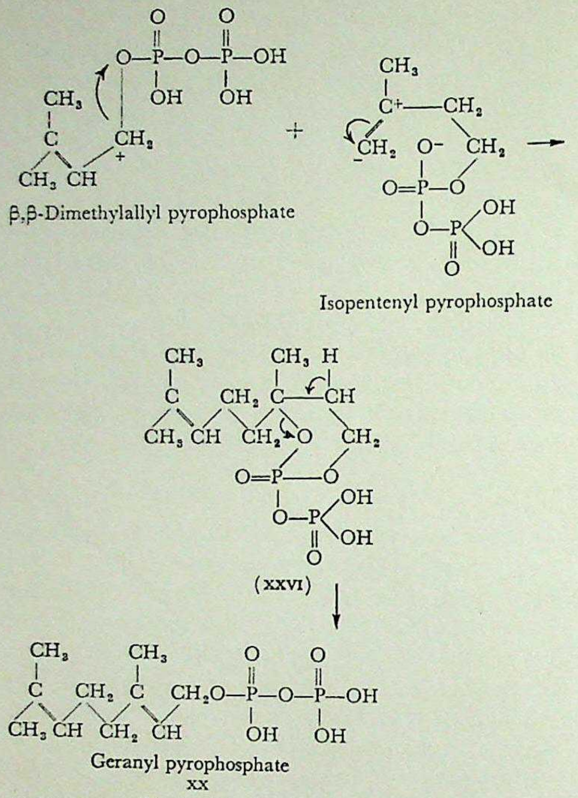


FIGURE 7 – *Mechanism for the dimerization of isopentenyl pyrophosphate to geranyl pyrophosphate* [38, 40–42].

the pathway. The existence of other phosphorylated intermediates of mevalonic acid was predicted [35] on the basis that ATP is required to convert mevalonic acid 5-phosphate to squalene. One of the new phosphorylated derivatives detected in this investigation was converted to squalene in the absence of ATP.

Two products from the enzymic conversion of mevalonic acid 5-phosphate by yeast in the presence of ATP have been isolated and identified [36, 37]. The first was identified as mevalonic acid 5-pyrophosphate (XVI). Radioactivity could be incorporated into this intermediate if either 1-^{14}C- or 2-^{14}C-mevalonic acid 5-phosphate was the starting material; it follows that conversion to the diphospho intermediate (XVI) proceeded with retention of the carboxyl group [37]. The second intermediate was identified as 3-methyl-3-butenyl 1-pyrophosphate (XVIII). As the formation of isopentenyl pyrophosphate from mevalonic acid 5-pyrophosphate requires ATP, phosphorylation must be an essential part of the elimination reaction. For the conversion of mevalonic acid 5-pyrophosphate to isopentenyl pyrophosphate, it was assumed [37] that phosphorylation occurs at the tertiary hydroxyl group, since the chemical lability of a tertiary ester would be expected to lead to unsaturation. The mechanism for the reaction leading from the triphospho intermediate (XVII) to isopentenyl pyrophosphate is visualized as a concerted decarboxylation-dephosphorylation [37]; protonation of the carbon chain was not included as a step in the mechanism, because earlier studies had shown that deuterium was not introduced during the biosynthesis of squalene in deuterium oxide.

The enzymic isomerization of isopentenyl pyrophosphate (XVIII) to β,β-dimethylallyl pyrophosphate (XIX) has been observed in aqueous extracts of yeast by F. Lynen and his colleagues [38, 39]. Geranyl pyrophosphate (XX) and farnesyl pyrophosphate (XXI) were also identified by these investigators as intermediates in the biosynthesis of squalene from mevalonic acid by cell-free yeast extracts. A mechanism, suggested by Bloch, Cornforth, and Lynen, for the synthesis of geranyl pyrophosphate from isopentenyl pyrophosphate is visualized (figure 7) as the alkylation of an isolated double bond by an allylic carbonium ion [38, 40, 41]. Isopentenyl pyrophosphate isomerizes to β,β-dimethylallyl pyrophosphate and probably exists as an equilibrium mixture of the two forms. One of the isomers can be regarded as the electrophile and the other as the nucleophile necessary for the condensation of two molecules to yield a dimer. The unsaturated moiety of isopentenyl

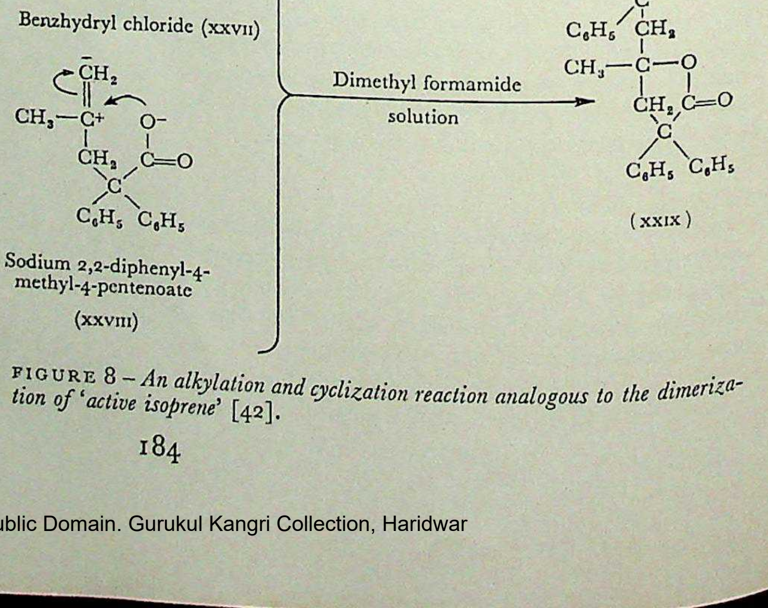

Benzhydryl chloride (XXVII)
Sodium 2,2-diphenyl-4-methyl-4-pentenoate
(XXVIII)
Dimethyl formamide solution
(XXIX)

FIGURE 8 – *An alkylation and cyclization reaction analogous to the dimerization of 'active isoprene'* [42].

pyrophosphate provides the nucleophilic centre; β,β-dimethylallyl pyrophosphate, on the basis of analogy to the behaviour of benzyl and other allyl esters, is susceptible to attack at the oxygenated carbon atom by the nucleophile.

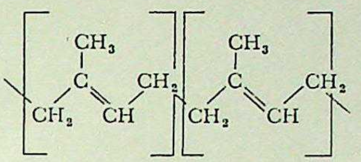
W. S. Johnson and R. A. Bell [42] have suggested that the alkylation reaction proceeds with the anchimeric assistance of the pyrophosphate moiety of isopentenyl pyrophosphate, leading to the cyclic intermediate (xxvi). The cyclic tertiary phosphate moiety should be highly unstable and should collapse readily to yield geranyl pyrophosphate. Support for the existence of such an intermediate was derived from a study of the alkylation reaction shown in figure 8. Isolation of the γ-lactone (xxix) of the predictable alkylation product not only supports the concept of a cyclic tertiary phosphate being formed during the polymerization of isopentenyl pyrophosphate, but is the basis for a new approach to cyclization reactions under neutral, or even alkaline, conditions.

The biosynthetic pathway from mevalonic acid to squalene in rat liver is similar to that in yeast cells (figure 5). According to G. Popják and his colleagues [43–46], the sequence includes such reactions as the conversion of mevalonic acid to mevalonic acid 5-phosphate, mevalonic acid 5-pyrophosphate, and isopentenyl pyrophosphate (xviii). The sequence continues with the isomerization of the last of these to dimethylallyl pyrophosphate (xix) and the condensation of these isomers to yield geranyl pyrophosphate (xx), which condenses with isopentenyl pyrophosphate to yield farnesyl pyrophosphate.

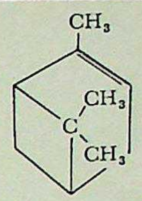
Lanosterol

Once it had been established that squalene was an intermediate in the biogenesis of cholesterol, R. B. Woodward and K. Bloch [47] proposed that lanosterol (xxiv) was intermediate between squalene and cholesterol on the biosynthetic pathway. The biosynthesis of lanosterol from labelled acetate in a tissue homogenate was demonstrated [48], and the further conversion of this labelled lanosterol to cholesterol was shown.

The cyclization of squalene to lanosterol is regarded as a single reaction involving the concerted participation of at least sixteen centres. A general mechanism proposed by A. Eschenmoser, L. Ruzicka, O. Jeger, and D. Arigoni [49] visualizes the formation of an intermediate carbonium ion after the reaction of a cationic activator, such as HO$^+$, with a properly folded substrate; for the conversion of squalene to lanosterol the reaction may be visualized as a series of nucleophilic attacks on electron-deficient centres by the electrons of double bonds. With a suitably folded squalene molecule, the concerted shift of electrons initiated by HO$^+$ is transmitted through five double bonds to form the tetracyclic structure (xxiii) containing a carbonium ion at C-20. After the 1,2-migration of two methyl groups and two hydrogen atoms, a proton is eliminated from the 9-position to form the stable lanosterol molecule.



Rubber
(xxx)



α-Pinene
(xxxi)

Cholesterol

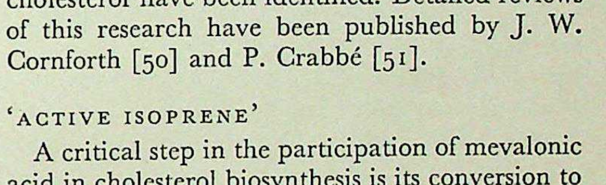
The conversion of lanosterol to cholesterol requires the oxidative cleavage of three methyl groups. Several intermediates and new precursors which function between lanosterol and cholesterol have been identified. Detailed reviews of this research have been published by J. W. Cornforth [50] and P. Crabbé [51].

'ACTIVE ISOPRENE'

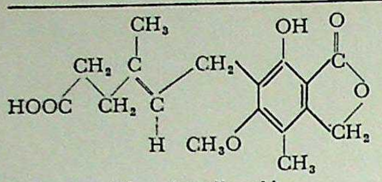
A critical step in the participation of mevalonic acid in cholesterol biosynthesis is its conversion to isopentenyl pyrophosphate (xviii), which is regarded as 'active isoprene'. This compound, with its isomer β,β-dimethylallyl pyrophosphate, constitutes the basic unit for biochemical polymerization. When polymers arising from this basic unit are subjected to other biochemical transformations, a wide variety of naturally occurring isoprenoid compounds and polymers result.

THE BIOSYNTHESIS OF ISOPRENOIDS

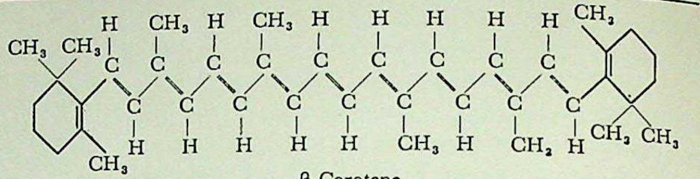
The importance of mevalonic acid in the biosynthesis of isoprenoid natural products has been substantiated by numerous investigations, based



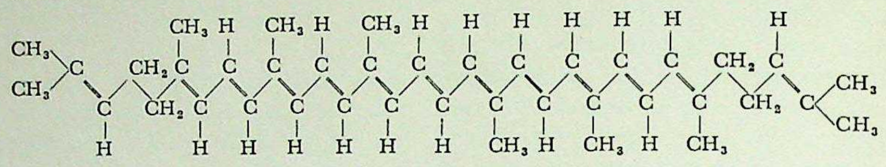
Mycelianamide
(xxxii)



Mycophenolic acid (xxxiii)



β-Carotene (xxxiv)

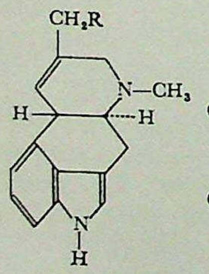


Lycopene (xxxv)

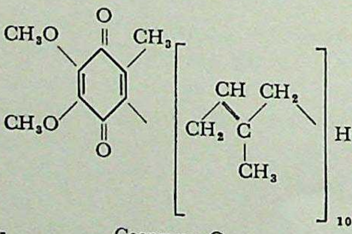
on the incorporation of radioactive mevalonic acid into a variety of molecules, such as rubber, terpenes, and alkaloids.

R. B. Park and J. Bonner [52] observed the incorporation of 2-^{14}C-mevalonic acid into rubber (xxx) by a crude enzyme system from latex of *Hevea brasiliensis*; the efficiency of conversion was about 2 per cent. Subsequent studies [53] showed a 9·3 per cent conversion of mevalonic acid to polyisoprene. The conversion of isopentenyl pyrophosphate into rubber has been demonstrated [54].

In the terpene field, R. G. Stanley [55] showed that seedlings of *Pinus attenuata* incorporate 2-^{14}C-mevalonic acid into the monoterpene fraction, which is composed almost entirely of α-pinene (xxxi). The incorporation of 2-^{14}C-mevalonic acid into the ten-carbon terpenoid side chain of mycelianamide (xxxii) by *Penicillium griseofulvum* and into the seven-carbon side chain of mycophenolic acid (xxxiii) by *Penicillium brevi-compactum* was also demonstrated [56]. The seven-carbon unit is regarded as being derived from a geranyl moiety.



Agroclavine (xxxvi, R = H) and elymoclavine (xxxvii, R = OH)



Coenzyme Q$_{10}$ (xxxviii)

The incorporation of 2-^{14}C-mevalonic acid into β-carotene (xxxiv) by growing cultures of the fungus *Phycomcyes blakesleeanus* has been demonstrated by G. D. Braithwaite and T. W. Goodwin [57], who also observed incorporation of the labelled factor into β-carotene by isolated slices of carrot root. It has been shown [58] that the mould *Mucor hiemalis* can incorporate 2-^{14}C-mevalonic acid into β-carotene. Enzymic synthesis of lycopene (xxxv) from 2-^{14}C-mevalonic acid in tomato homogenates has been observed [59].

The biosynthetic significance of mevalonic acid has been demonstrated in other fields too. Thus it has been shown [60] that *Claviceps purpurea*, the parasitic fungus of rye, incorporates ^{14}C isotope from 2-^{14}C-mevalolactone into the alkaloids agroclavine (xxxvi) and elymoclavine (xxxvii). This, and 4-^{14}C-mevalonic acid, can be incorporated into the isoprenoid side chain of coenzyme Q$_{10}$ (xxxviii) [61, 62].

Figures 2, 3, and 4 are reproduced from 'Advances in Enzymology', Volume XXIII, by permission of Interscience Publishers Inc.

REFERENCES

[1] NOVAK, A. F., HAUGE, S. M., and CARRICK, C. W. *Poult. Sci.*, **26**, 604, 1947.

[2] NOVAK, A. F. and HAUGE, S. M. *J. biol. Chem.*, **174**, 647, 1948.

[3] SKEGGS, H. R., WRIGHT, L. D., CRESSON, E. L., MACRAE, G. D. E., HOFFMAN, C. H., WOLF, D. E., and FOLKERS, K. *J. Bact.*, **72**, 519, 1956.

[4] OTT, W. H., DICKINSON, A. M., VAN INDERSTINE, A., BAZEMORE, A. W., PAGE, A. C., and FOLKERS, K. *J. Nutr.*, **64**, 525, 1958.

[5] TAMURA, G. *J. gen. appl. Microbiol.*, **2**, 431, 1956.

[6] TAMURA, G. and FOLKERS, K. *J. org. Chem.*, **23**, 772, 1958.

[7] WRIGHT, L. D., CRESSON, E. L., SKEGGS, H. R., MACRAE, G. D. E., HOFFMAN, C. H., WOLF, D. E., and FOLKERS, K. *J. Amer. chem. Soc.*, **78**, 5273, 1956.

[8] WOLF, D. E., HOFFMAN, C. H., ALDRICH, P. E., SKEGGS, H. R., WRIGHT, L. D., and FOLKERS, K. *Ibid.*, **78**, 4499, 1956.

[9] *Idem. Ibid.*, **79**, 1486, 1957.

[10] Hoffman, C. H., Wagner, A. F., Wilson, A. N., Walton, E., Shunk, C. H., Wolf, D. E., Holly, F. W., and Folkers, K. *Ibid.*, **79**, 2316, 1957.

[11] Cornforth, J. W., Cornforth, R. H., Popják, G., and Youhotsky Gore, I. *Biochem. J.*, **69**, 146, 1948.

[12] Cornforth, J. W., Cornforth, R. H., Pelter, A., Horning, M. G., and Popják, G. *Tetrahedron*, **5**, 311, 1959.

[13] Shunk, C. H., Linn, B. O., Huff, J. W., Gilfillan, J. L., Skeggs, H. R., and Folkers, K. *J. Amer. chem. Soc.*, **79**, 3294, 1957.

[14] Lynen, F. and Grassl, M. *Z. physiol. Chem.*, **313**, 291, 1958.

[15] Lynen, F., Eggerer, H., Henning, U., and Kessel, I. *Angew. Chem.*, **70**, 738, 1958.

[16] Henning, U., Möslein, E. M., and Lynen, F. *Arch. Biochem. Biophys.*, **83**, 259, 1959.

[17] Eberle, M. and Arigoni, D. *Helv. Chim. Acta*, **43**, 1508, 1960.

[18] Bloch, K. and Rittenberg, D. *J. biol. chem.*, **143**, 297, 1942; **145**, 625, 1942.

[19] Little, H. N. and Bloch, K. *J. biol. Chem.*, **183**, 33, 1950.

[20] Cornforth, J. W., Hunter, G. D., and Popják, G. *Biochem. J.*, **54**, 597, 1953; *Arch. Biochem. Biophys.*, **42**, 482, 1953.

[21] Wuersch, J., Huang, R. L., and Bloch, K. *J. biol. Chem.*, **195**, 439, 1952.

[22] Bonner, J. and Arreguin, B. *Arch. Biochem. Biophys.*, **21**, 109, 1949.

[23] Brady, R. O. and Gurin, S. *J. biol. Chem.*, **189**, 371, 1951.

[24] Rabinowitz, J. L. and Gurin, S. *Ibid.*, **208**, 307, 1954.

[25] Rudney, H. *J. Amer. chem. Soc.*, **76**, 2595, 1954; **77**, 1698, 1955.

[26] *Idem. J. biol. Chem.*, **227**, 363, 1957.

[27] Tavormina, P. A., Gibbs, M. H., and Huff, J. W. *J. Amer. chem. Soc.*, **78**, 4498, 1956.

[28] Tavormina, P. A. and Gibbs, M. H. *Ibid.*, **78**, 6210, 1956.

[29] Eggerer, H. and Lynen, F. *Liebigs Ann.*, **608**, 71, 1956.

[30] Folkers, K., Shunk, C. H., Linn, B. O., Robinson, F. M., Wittreich, P. E., Huff, J. W., Gilfillan, J. L., and Skeggs, H. R. 'Ciba Foundation Symposium on Biosynthesis of Terpenes and Sterols', p. 20. J. and A. Churchill Ltd, London. 1959.

[31] Langdon, R. G. and Bloch, K. *J. Amer. chem. Soc.*, **74**, 1869, 1952.

[32] Cornforth, J. W. and Popják, G. *Biochem. J.*, **58**, 403, 1954.

[33] Admur, B. H., Rilling, H., and Bloch, K. *J. Amer. chem. Soc.*, **79**, 2646, 1957.

[34] Tchen, T. T. *Ibid.*, **79**, 6344, 1957; *J. biol. Chem.*, **233**, 1100, 1958.

[35] Phillips, A. H., Tchen, T. T., and Bloch, K. *Fed. Proc.*, **17**, 289, 1958.

[36] Chaykin, S., Law, J., Phillips, A. H., Tchen, T. T., and Bloch, K. *Proc. Nat. Acad. Sci.*, **44**, 998, 1958.

[37] Bloch, K., Chaykin, S., Phillips, A. H., and de Waard, A. *J. biol. Chem.*, **234**, 2595, 1959.

[38] Lynen, F., Eggerer, H., Henning, U., and Kessel, I. *Angew. Chem.*, **70**, 738, 1958.

[39] Lynen, F., Agranoff, B. W., Eggerer, H., Henning, U., and Möslein, E. M. *Angew. Chem.*, **71**, 657, 1959.

[40] Rilling, H. C. and Bloch, K. *J. biol. Chem.*, **234**, 1424, 1959.

[41] Cornforth, J. W. *Tetrahedron Lett.*, No. 19, 29, 1959.

[42] Johnson, W. S. and Bell, R. A. *Ibid.*, No. 12, 27, 1960.

[43] de Waard, A. and Popják, G. *Biochem. J.*, **73**, 410, 1959.

[44] Levy, H. R. and Popják, G. *Ibid.*, **72**, 35P, 1959; **75**, 417, 1960.

[45] Popják, G. *Tetrahedron Lett.*, No. 19, 19, 1959.

[46] Goodman, de W. S. and Popják, G. *Biochem. J.*, **74**, 35P, 1960; *J. Lipid Res.*, **1**, 286, 1960.

[47] Woodward, R. B. and Bloch, K. *J. Amer. chem. Soc.*, **75**, 2023, 1953.

[48] Clayton, R. B. and Bloch, K. *J. biol. Chem.*, **218**, 305, 319, 1956.

[49] Eschenmoser, A., Ruzicka, L., Jeger, O., and Arigoni, D. *Helv. Chim. Acta*, **38**, 1890, 1955.

[50] Cornforth, J. W. *J. Lipid Res.*, **1**, 3, 1959.

[51] Crabbé, P. *Rec. chem. Progr.*, **20**, 189, 1959.

[52] Park, R. B. and Bonner, J. *J. biol. Chem.*, **233**, 340, 1958.

[53] Kekwick, R. G. O., Archer, B. L., Barnard, D., Higgins, G. M. C., Mc Sweeney, G. P., and Moore, C. G. *Nature, Lond.*, **184**, 268, 1959.

[54] Henning, U., Möslein, E. M., and Lynen, F. *Biochem. Z.*, **333**, 534, 1961.

[55] Stanley, R. G. *Nature, Lond.*, **182**, 738, 1958.

[56] Birch, A. J., English, R. J., Massy-Westropp, R. A., and Smith, H. *J. chem. Soc.*, 369, 1958.

[57] Braithwaite, G. D. and Goodwin, T. W. *Biochem. J.*, **66**, 13P, 31P, 1957.

[58] Grob, E. C. *Chimia*, **11**, 338, 1957.

[59] Shneour, E. and Zabin, I. *J. biol. Chem.*, **234**, 770, 1959.

[60] Birch, A. J., McLoughlin, B. J., and Smith, H. *Tetrahedron Lett.*, No. 7, 1, 1960.

[61] Gloor, U. and Wiss, O. *Experientia*, **14**, 410, 1958.

[62] Gloor, U., Schindler, O., and Wiss, O. *Helv. Chim. Acta*, **43**, 2089, 1960.

The Mohole

SIR EDWARD BULLARD

It has long been known that there is a change in the properties of the rocks of the earth at a depth of about 35 km under the continents. More recently it has been shown that this discontinuity, called the Moho, lies at a depth of only 5 km below the ocean floor. This article describes the plans to reach this discontinuity by drilling a hole, known as the Mohole, and discusses the results that have been obtained by drilling a trial hole to a lesser depth.

THE MOHOROVIČIĆ DISCONTINUITY

On the morning of the 8th October 1909, an earthquake occurred about 40 km south of Zagreb, in Croatia. Readable records of this earthquake were obtained at stations all over Europe and were examined by Dr Andrija Mohorovičić, the director of the Meteorological Observatory at Zagreb [5]. He found that, as with other earthquakes, there were two kinds of waves present, which could be interpreted as the compressional and distortional waves required by the theory of the propagation of elastic waves through a solid. In compressional or P waves, the oscillatory motion of the particles takes place along a line in the direction in which the wave is travelling, as is shown in figure 1a. In distortional or S waves, the motion of the particles is at right angles to the direction of travel of the wave (figure 1b). The existence and significance of these two types of wave had been known for some years; the new discovery made by Mohorovičić was that there are two P waves. At short distances, the first wave to arrive travels at a speed that varies from place to place within the range $5\frac{1}{2}$–$6\frac{1}{2}$ km/s. At about 170 km this is overtaken by another P wave travelling at 8·1 km/s. Over a range of distance on each side of this, usually out to about 800 km, both waves can be recorded, as is shown in the seismogram of figure 5. Beyond this the slower wave fades out. The travel times of the two waves are shown as a function of distance in figure 2.

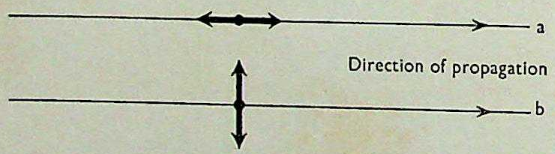


FIGURE 1 – *Motion of a particle in* (a) *a compressional wave and* (b) *a distortional wave.*

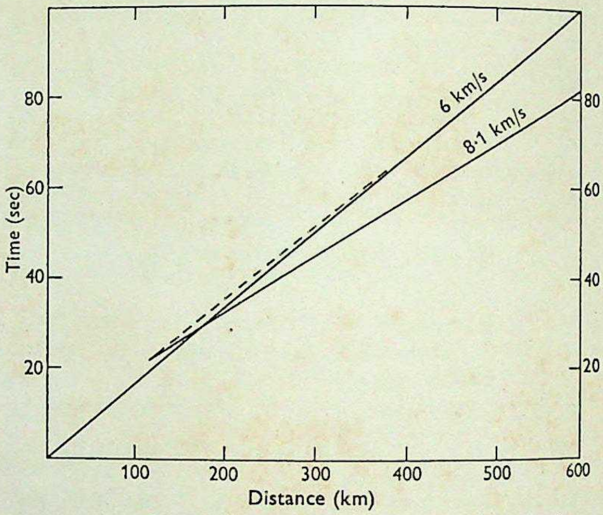


FIGURE 2 – *Travel times of seismic waves. The broken line gives the travel time of the reflected wave.*

Mohorovičić suggested that the slower of the two P waves travels by the direct route between the focus of the earthquake and the seismograph, and that the faster wave is refracted at a discontinuity at a depth of about 50 km, as is shown in figure 3. This suggestion has been very widely accepted, and the discontinuity is known as the Mohorovičić discontinuity, or Moho. Mohorovičić overestimated its depth, which varies, from place to place, between 30 and 45 km. If the discontinuity is sharp, it should be possible to observe a wave reflected from the Moho; the path of this wave is shown by a broken line in figure 3. Several workers claim to have detected this reflected wave, but it is difficult to be sure of its existence, because of the continuous motion shown on the record

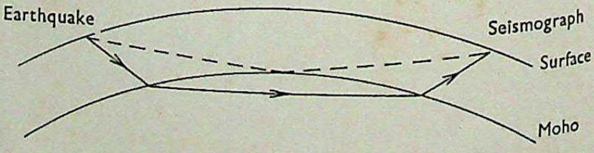


FIGURE 3 – *Section of the Earth, showing the refraction of a seismic wave at the Moho. The broken line shows the track of the reflected wave. The curvature of the Earth is exaggerated.*

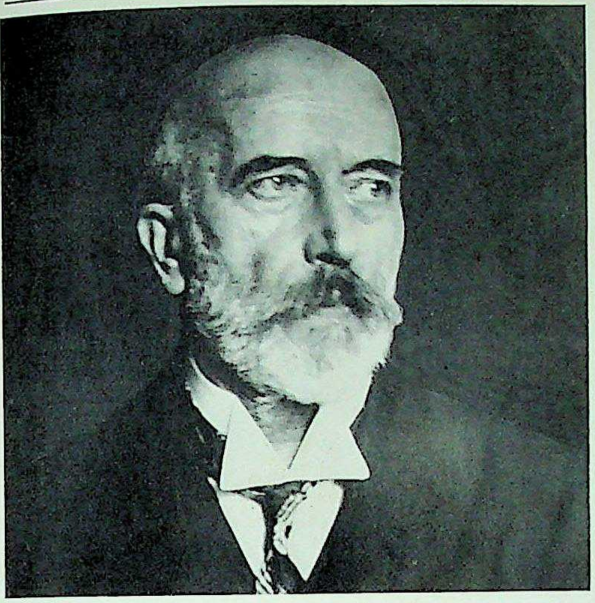


FIGURE 4 – *A. Mohorovičić.*

following the P waves. The matter is of some importance, since a reflection is only possible if the change in seismic velocity occurs over a range of depth of less than a wavelength. Unless the reflected waves can be observed it is difficult to be sure whether the Moho is a true discontinuity or whether the change in seismic velocity is spread over a range of depth of a few kilometres.

Observations at sea, using explosions as a source of energy and hydrophones in the water as detectors, have shown a similar discontinuity at a depth of about 5 km beneath the floor of the deep ocean, that is about 10 km beneath the sea surface. It is difficult to trace the Moho through the transition region between a continent and an ocean. The best-investigated region is that off the eastern coast of the United States [3], a generalized section of which is shown in figure 11. There is little doubt that the continental and oceanic Moho are connected in the way shown in the figure. The velocity

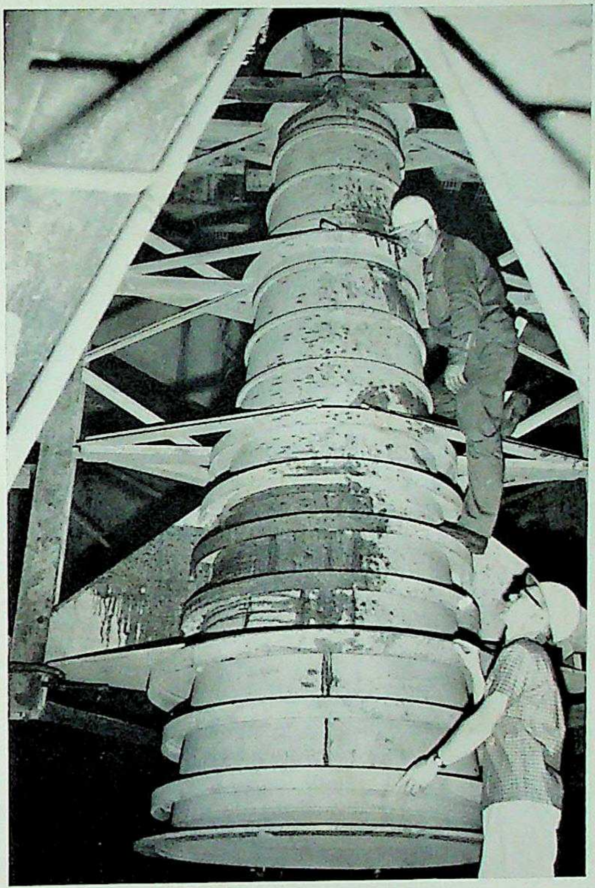


FIGURE 6 – *Conical guide for drill pipe. (National Science Foundation photo.)*

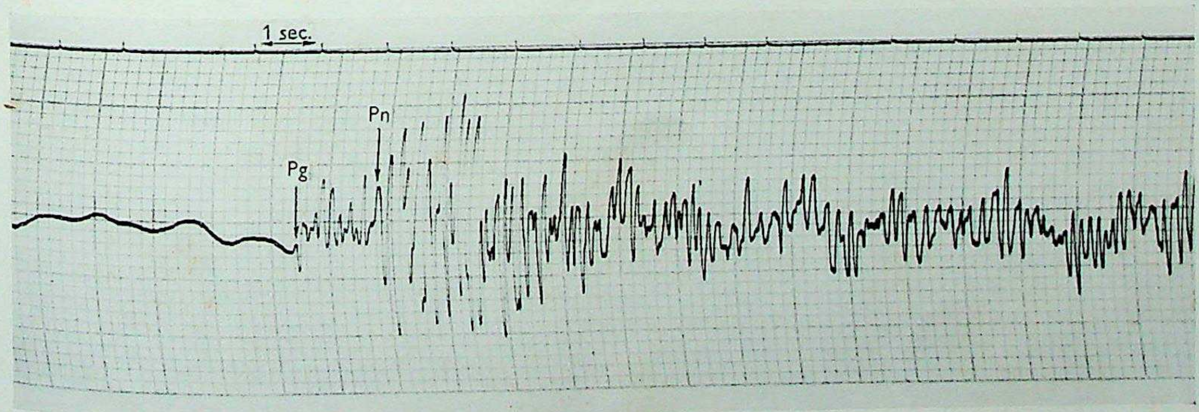


FIGURE 5 – *Seismogram taken at a distance of 113 km from an explosion (approx. 500 kg Torpex) in Loch Striven, near Dunoon, taken by the seismic group of the Atomic Energy Authority. Pg is the arrival of the direct wave and Pn that of the wave refracted at the Moho.*

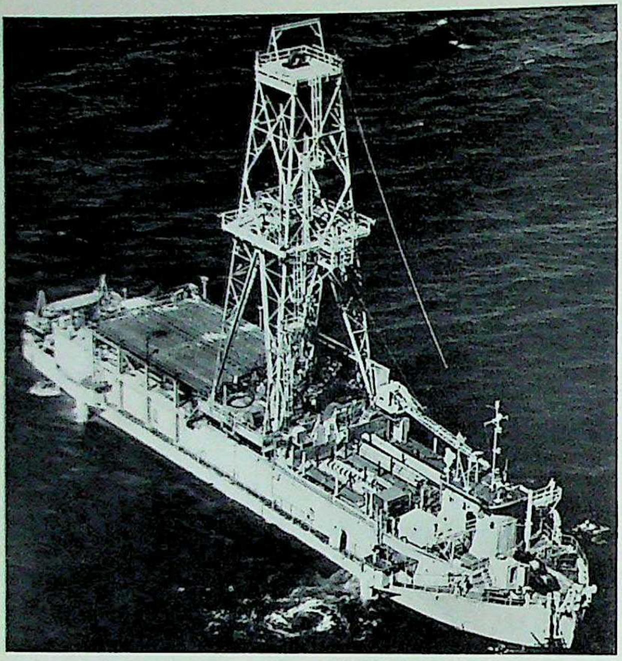


FIGURE 7 – CUSS I.

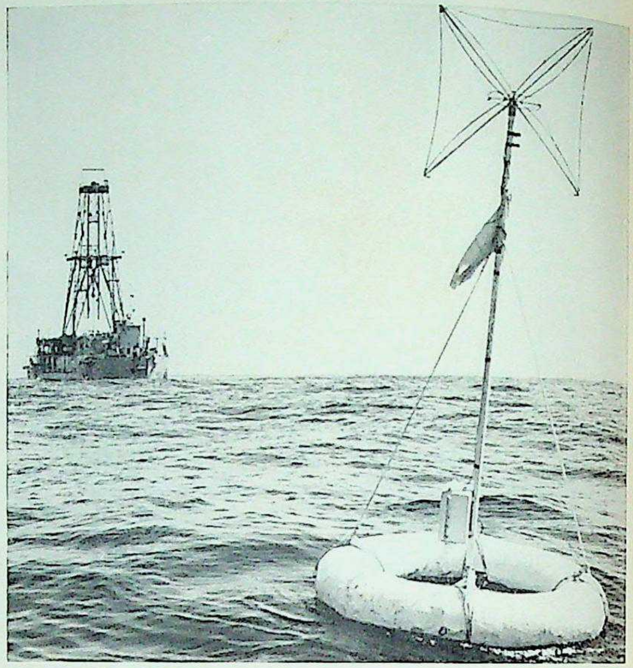


FIGURE 8 – CUSS I *and float with radar reflector.* (*This figure and figure* 10 *reproduced by courtesy of Global Marine Exploration Company.*)

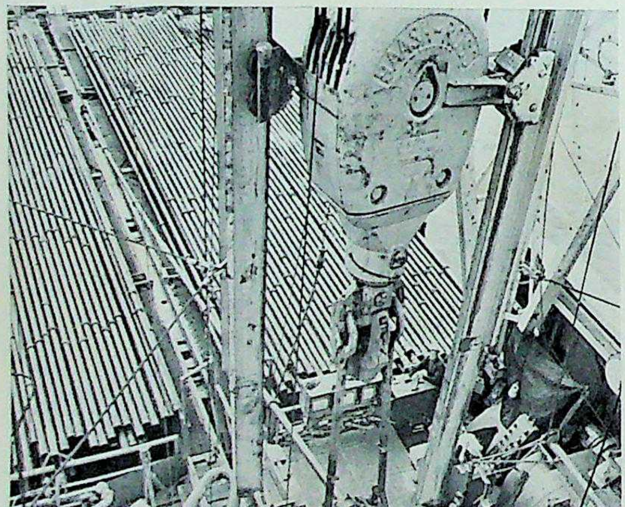


FIGURE 9 – *Raising an* 18 *m length of drill pipe.*

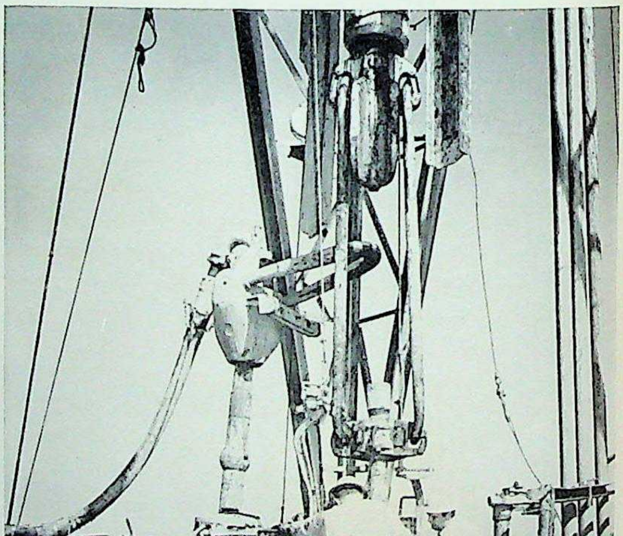


FIGURE 10 – *The rotating table. The drill pipe hangs from the block, the lower part of which is at the top of the figure. The fitting through which sea water is pumped down the drill pipe can be seen on the left.*

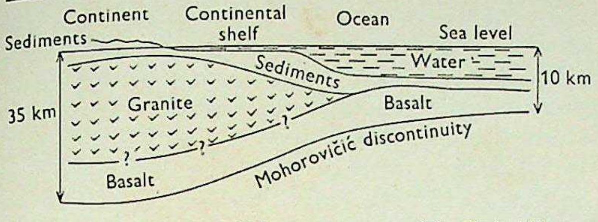


FIGURE 11 – *Section off the eastern coast of the United States.*

below the oceanic discontinuity is usually within 0·2 km/s of the continental average of 8·1 km/s.

Above the discontinuity the continental rocks appear to be different from the oceanic ones. Beneath the sediments of the continents there lie rocks similar to those exposed in the ancient shield areas, such as Canada and east Africa. These rocks are typically granites, gneisses, and schists, and probably represent the worn-down roots of ancient mountains. The seismic velocity above the Moho often increases with depth, probably because of an increase in the proportion of basic rocks. In some places there may be a separate layer of more basic rocks between the granite and the Moho. Beneath the oceans there is a variable thickness of soft sediments in which seismic waves travel a little faster than they do in water. The thickness of this material is normally between 0·1 and 1 km. Beneath this comes 'layer 2' with a seismic velocity in the neighbourhood of 4·8 km/s, and a thickness of about 2 km. The 3 km between this and the Moho is called 'layer 3', with a seismic velocity of 6½–7 km/s. Since there is an enormous preponderance of basalt among oceanic igneous rocks, it is usually supposed that both layer 2 and layer 3 are composed of basalt, perhaps interlayered with consolidated sediments.

The structure shown in figure 11 is based primarily on the seismic results and on petrological probabilities; it is, however, confirmed by measurements of the acceleration due to gravity. On the average, gravity at sea level over a continent is within a few parts in a million of that over the neighbouring ocean in the same latitude. From this it follows that the mass beneath unit area is the same for the continents and the oceans. This suggests that the material beneath the Moho has a higher density than that above, and that the greater amount of this heavy material beneath the ocean compensates for the lightness of the sea-water relative to the near-surface rocks of the continents. The density difference required is about 0·5 g/cm^3, which suggests a density of about 3·2 g/cm^3 beneath the Moho. This is an example of the principle of isostasy, which was discovered in the middle of the last century by Archdeacon Pratt of Calcutta, from an examination of the deflections of the vertical observed by the Survey of India. The principle of isostasy states that the Earth's topographic features float in hydrostatic equilibrium in the same way as an iceberg floats in the sea. It is true in a general way for most large features, but applies to averages over considerable areas rather than to individual points. The compensation for the visible masses of topographic irregularities seems to be achieved by variations in the depth of the Moho, and there is a strong correlation of this depth with the height of the topography and with the Bouguer gravity anomaly, as is shown in figure 12. (The Bouguer anomaly represents the attraction or lack of attraction of hidden differences in density after allowance has been made for the visible topography.)

THE MATERIAL BENEATH THE MOHO

The Mohorovičić discontinuity appears to be one of the major features of the earth. It occurs beneath all the continents and all the oceans. Above it there is the infinite variety and complexity of geology; but below it, if we may judge from the constancy of the seismic velocities, there is a much greater uniformity.

In spite of its obvious importance, very little is known about the Moho. The seismic velocities above and below it are known and we have some idea of the density jump across it, but we do not know whether it represents a sharp contact between two different materials or whether the change from one to the other is spread over a range of depth of a few kilometres.

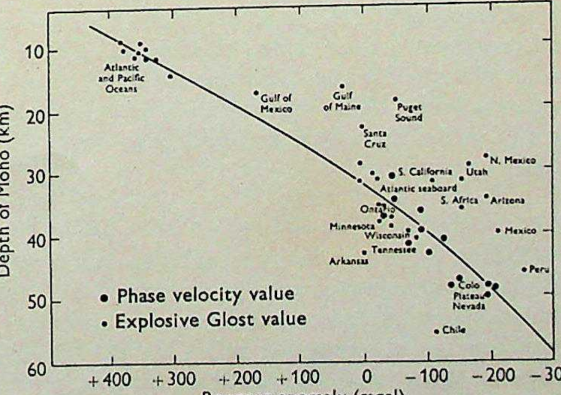


FIGURE 12 – *Correlation between the depth of the Moho and the Bouguer anomaly. The results shown by the large dots were obtained by the method described in the text, those shown by the small dots by a method using surface waves from earthquakes (from Woollard [6]).*

Petrology gives some hints as to the nature of the material. There are, in various parts of the world, intrusive rocks containing minerals that can be formed only at high pressures. The best known of these are the diamond pipes of South Africa. The stability relations of diamond and graphite are now well understood, and it is clear that the diamonds must have been formed below the Moho. The associated rocks must, therefore, also come from great depths, and must represent a more or less contaminated sample of the material beneath the Moho. The rocks of the diamond pipes are mineralogically complicated, but consist predominantly of iron and magnesium silicates, particularly olivine, garnet, and enstatite. On the basis of these and other occurrences of rocks derived from great depths it has often been suggested that the material beneath the Moho is either peridotite, which is essentially composed of magnesium-rich olivine, $(Mg,Fe)_2SiO_4$, or eclogite, a rock composed of garnets and pyroxenes.

The comparison of the densities, and elastic wave velocities, of proposed materials with the properties of those actually occurring below the Moho is complicated by the changes in crystal structure found in many silicates at about the pressures and temperatures of the continental Moho. In fact it has been suggested that the discontinuity might be due to phase changes in basalt without any change of material [2, 4]. This is quite a plausible view for the Moho beneath the continents, but can scarcely be correct beneath the oceans. The pressure at the oceanic Moho is only about 1800 kg/cm^2 and the temperature 120° C; a thoroughgoing reorganization of the silicate lattices at so low a temperature and pressure seems improbable. If, as seems likely, the oceanic and continental Mohos are continuous, it is probable that the continental Moho also is not due to a phase change. The expectation is that the Moho is underlain by a material different from that above.

THE MOHOLE PROJECT

Clearly there are many possibilities, and it is unlikely that the nature of the material beneath the Moho can be determined by laboratory experiment or by observations taken at the Earth's surface. The most satisfactory solution of the difficulty would clearly be to drill a hole and obtain samples. The deepest hole so far drilled on land reached a depth of 7·7 km; there seems no prospect of improving techniques sufficiently to drill to the depth of 35 km needed to reach the Moho on the continents. At sea, where the Moho often lies beneath 5 km of water and 5 km of sediments, the prospects appear more hopeful, though there are still formidable difficulties. The present attempt stems from a proposal by H. Hess and W. Munk in 1957 to the American Miscellaneous Society (AMSOC). This society originated in 1952 as an informal and not wholly serious association of American earth scientists, formed 'to encourage miscellaneous projects'. In April 1957, the society formed a deep-drilling committee, with G. Lill as chairman, to consider the possibility of drilling to the Moho. In September of the same year, the triannual meeting of the International Union of Geodesy and Geophysics took place in Toronto. The matter was not on the agenda, but a good deal of private discussion took place and the possibilities received publicity in the press. In spite of considerable scepticism, the Associations of Oceanography and of Seismology were persuaded to adopt resolutions supporting the project.

At this stage it was clear that there was a case good enough to require investigation. As AMSOC had no officers, no defined membership, and no legal existence, its deep-drilling committee was taken over by the National Academy of Sciences, which received a small grant from the National Science Foundation to study ways and means. W. Bascom was appointed director of the project. He recruited the necessary staff and enlisted the help of drilling contractors, oil companies, and naval architects. The hole was christened the 'Mohole', a convenient and euphonious abbreviation of 'the hole to reach the Mohorovičić discontinuity'.

The feasibility study was concerned with four main topics:

1. How can a ship be kept within a hundred metres of a given point in the open ocean?
2. How can 10 km of drill stem be handled at sea and used to drill a hole to a depth below the sea surface greater than that ever achieved on land?
3. How can the drill be re-inserted in the hole after withdrawing it to change a bit?
4. Where should the hole be drilled?

In the summer of 1960 Bascom issued a report in which all these questions were discussed and no insuperable difficulties found.

DRILLING IN THE DEEP SEA

The next stage was to gain practical experience, to define the difficulties more clearly, and to demonstrate the feasibility of the project by actually drilling a short distance into the floor of the deep ocean. This was done in April 1961. Most drilling in shallow water has been done from derricks supported on platforms resting on the sea floor. This is not a method that can be used in the deep ocean, most of which has a depth between $3\frac{1}{2}$ and 6 km. In such deep water it is necessary to operate from a floating platform. Drilling rigs have been mounted on various kinds of ships and barges and have been used successfully at sea, particularly off the coast of California. For the preliminary trials for the Mohole the AMSOC committee selected the barge *CUSS I*. The name of this vessel is formed from the initial letters of the names of her owners, the Continental, Union, Shell and Superior Oil Companies.

CUSS I is a converted naval freight barge; she is 79 m long and 15 m wide, and draws 4·6 m when loaded. She carries a derrick 30 m high which drills through a hole in the centre of the ship (see figures 7 and 8). The trial hole was drilled between Guadalupe Island and the west coast of Mexico at a point about 74 km from the island and 220 km from shore (117°30′ W, 29°0′ N); the depth of the water was 3760 m.

When drilling at the test site, *CUSS I* had some 3700 m of drill pipe between her and the ocean floor. If the ship moves, this pipe is bent, and may fracture if the movement is too great. The amount of movement needed to break the pipe is not known, but is believed, from calculation and from experience in shallow water, to be about 200 m. It is therefore desirable to keep the ship within say 60 m of a point directly above the hole being drilled. It is impracticable, or at any rate very expensive, to anchor a large ship with this accuracy in deep water; the varying pulls of wind and current would cause the anchor cables to tighten on one side and to sag on the other, and this would allow the ship to move through some hundreds of metres in an unpredictable manner. It was therefore decided to hold the barge in position, without moorings, by manœuvring her with the aid of four screws, one at each corner. Each of these screws could be rotated about a vertical axis, so that it could exert its thrust in any direction. The combination of the four, therefore, enabled the ship to be moved forward, backwards, or sideways and to be rotated. The speed and alignment of the screws were controlled by a single handle. This was connected to the engines and screws in such a way that the ship moved in the direction in which the handle was pushed. An outer ring round the handle controlled the direction of the ship's head.

These arrangements gave a very convenient and sensitive control of the ship's position and orientation. The marks relative to which the ship was held stationary were a ring of floats attached to totally submerged moored buoys 50 m below the surface (figure 13). A suitable buoy can be moored on a 6 mm steel wire, which it will hold very close to the vertical. The surface float (figure 8) experiences only small forces from wind and current and can be kept within 30 m of the point over the anchor weight. The position of the ship in a ring of buoys is determined by radar and by an acoustic method. Each float carries a radar reflector, which enables its position relative to the ship to be displayed on a 'plan position indicator' visible to the steersman. He sees each float as a bright spot of light on the radar screen and endeavours to keep these spots in marked positions. A similar display is produced by an acoustic system using transponders attached to the buoys. There are thus two entirely separate methods of positioning the ship, both having more than the required accuracy. The whole system worked very well and the ship was kept within 60 m of the required position for many days, in spite of winds of 50 km/h and 4 m waves. This aspect of the problem appears to have found an entirely satisfactory solution, though it is probable that some form of cycloidal propeller will replace the four outboard motors in the larger ship that will be needed to drill the Mohole. The steersman will probably be replaced by an automatic device, in view of the many months of operation that will be needed.

The trial drilling was done with diamond bits with an axial hole. For most of the drilling these were attached to an orthodox drilling rig; the bit was at the lower end of a string of drill pipes 14 cm in diameter screwed together from 9 m sections. 18 m lengths were stowed in racks and raised as required by a pulley block. This block was restrained by guides, which can be seen in figure 9; these are not fitted in a rig for drilling on land, but are necessary at sea to prevent the block from swinging from side to side as the barge rolls.

As the barge rolls and pitches the drill pipe is bent. Calculation shows that the maximum strain occurs near the surface and that, if no precautions were taken, the maximum allowable roll or pitch would be $\frac{1}{2}$°. This will be exceeded in quite moderate weather. The difficulty was overcome by providing a conical guide round the uppermost 16 m of the drill pipe (figures 6 and 14). The guide causes the bend near the upper end of the pipe to be spread over a greater length than if the

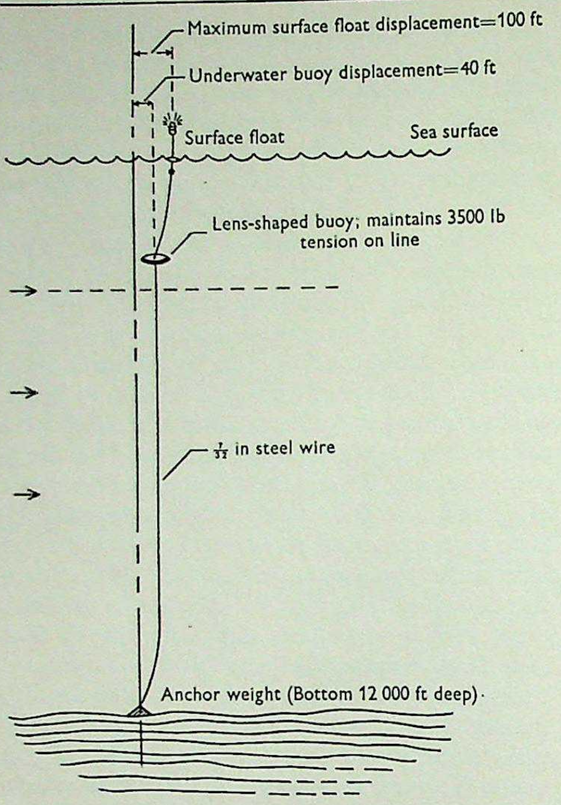

Maximum surface float displacement=100 ft
Underwater buoy displacement=40 ft
Surface float
Sea surface
Lens-shaped buoy; maintains 3500 lb tension on line
$\frac{1}{32}$ in steel wire
Anchor weight (Bottom 12 000 ft deep)

FIGURE 13 – *A marker buoy.* (*This figure and figure* 14 *by courtesy of Weidenfeld and Nicolson Ltd, publishers in Great Britain of 'A hole in the bottom of the sea', by W. Bascom* [1].)

guide had not been present, as is shown in figure 14. With the guide a roll of 6½° can be tolerated.

Chips of rock were removed from the hole by pumping sea water down the drill pipe and letting it flow through the bit, up the hole again, and out at the sea floor. The fitting for the top of the pipe can be seen in figure 10. The bit ground away a hollow cylinder of rock, leaving a central core in a tube on the axis. This tube and the enclosed core of rock could be recovered by pulling them up the drill pipe by a wire (the sandline). The tube could be replaced by carrying it down with the stream of sea water pumped into the drill pipe. It was thus possible to take many lengths of core without withdrawing the drill pipe and bit. When the bit became blunt or some accident prevented further progress, the hole was abandoned and a new hole started. In drilling a deeper hole it would be necessary to be able to re-enter a hole with a new bit.

Some tests were also made with a turbo drill. In this the bit was rotated by a turbine at the bottom of the hole, driven by the stream of sea water pumped down the drill pipe. This method has the advantage that it is not necessary to rotate the drill pipe, and the very large frictional forces between the pipe and the hole are thus avoided. It is also preferable to orthodox rotary drilling in that the diamond bit can be rotated faster, which leads to quicker and more economical drilling.

THE RESULTS OF THE TRIAL HOLES

Three trial holes were drilled close together at the site near Guadalupe Island. The primary purpose of making these holes was to develop methods of drilling in deep water, to find where the problems lay, and to establish the practicability of going deeper. A considerable amount of scientific information was, however, obtained from these holes. The sediments were found to be 180 m in thickness and to be underlain by basalt. The thickness agrees with measurements by explosion seismology, and the finding of basalt is no surprise, since it is by far the commonest igneous rock in the oceans and forms the great volcanic pile of Guadalupe Island only 74 km away. The surprising thing is that the whole 180 m of sediments appears to be of Miocene age, though at the time of writing only a preliminary examination of the core has been made. It is virtually certain that the sea covered this site long before the Miocene; in fact, there is no reason to suppose that it has not been an oceanic area for all past time. It seems highly improbable that the Miocene is the only period in which sediment was deposited, and we must therefore conclude that other sediments are concealed beneath the basalts of the Pacific. There is nothing unreasonable or surprising about this, but there had been no evidence for it previously; it does ease the problem posed

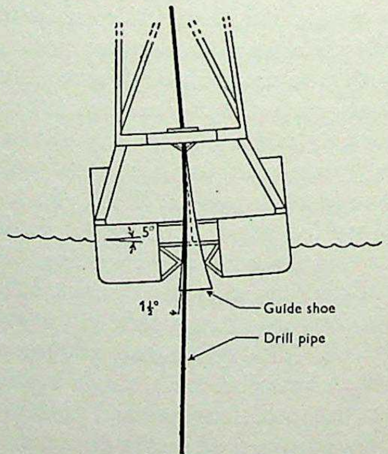


FIGURE 14 – *The effect of rolling with drill pipe in conical guide.*

by the small thicknesses of unconsolidated sediment found all over the world by the seismic method. We can now expect that much of the material between the Moho and the ocean floor will be found to be composed of sediments interlayered with basalt. The total length of time represented by the Miocene period is usually estimated at 14 million years; the sediments penetrated represent less than the full span of the period. If they are assumed to have been deposited in 5 million years, the rate of sedimentation is 4 cm/1000 yr, which is several times the average rate determined from short cores. Why sediment should have accumulated rapidly at this point during the Miocene and very slowly, or not at all, in the recent past, is not known.

The rock beneath the sediment is a fresh-looking crystalline basalt with a millimetre or two of glass at its upper surface. So far as the writer could see, there was no baked sediment sticking to it; it is therefore to be supposed that it flowed out on the ocean floor and was not intruded as a sill into pre-existing sediment. The basalt was penetrated for 15 m and a large amount of core collected. The results of the examination of this material will be awaited with interest. It will presumably be possible to determine its age, as well as its chemical composition and physical properties.

Temperature measurements in one of the holes gave a heat flow about twice the average for the oceans, this high value being in agreement with measurements made in the vicinity with short probes. These measurements give a most welcome confirmation of the results from the probes, since, in spite of the care with which possible errors in the probes have been investigated, it was not possible to feel entirely certain that measurements over a range of depth of a few metres gave really significant temperature gradients.

PROBLEMS FOR THE FUTURE

After the successful completion of the trial holes we may expect future work to divide into two streams. First, with the existing technique or slight extensions of it, and with ships the size of *CUSS I*, holes up to 300 m can certainly be drilled into the material beneath the floor of most parts of the ocean. This opens the possibility of studying in detail the sediments and lavas to that depth and of following the history of the oceans for a very considerable period in the same way that it has been followed on land. This should lead in a few years to the solution of some of the major problems of geology. It will be possible to establish beyond doubt whether the land masses that have been supposed to exist off many continental coasts have foundered in the present oceans; it is the expectation of the writer that such submerged continents will not be found. It should also be possible to find the age of each ocean basin and, from changes in the sediments, to study past changes in the distribution of land and sea, and perhaps changes of climate and ocean currents. The cost of such a programme would not be high. *CUSS I* cost about three million dollars, and it would probably be possible to operate a similar ship for something under one million dollars per year. It would seem well worth building at least one such ship instead of one or more of the proposed new conventional oceanographic ships.

The other line of development will be the attempt to reach the Moho. Since so deep a hole cannot be drilled with a single bit, it will be necessary to withdraw the drill pipe and bit, and re-enter the same hole with a new bit. This is an operation that has not been tried in deep water. There are at least two possible methods. A pair of wires might run from the ship to the bottom carrying a horizontal bar between them. If the lower end of the drill pipe were attached to the centre of this bar, it could slide down the wires and direct the bit into the hole. This method has been used in shallow water. If it were used in deep water, there is some danger of the wires and the drill pipe becoming entangled. Also it would not provide a return path to the ship for fluid pumped down the hole, and the drilling fluid would therefore have to be sea water, as in the trial holes. In drilling on land it is usual to use a thixotropic mud as a drilling fluid, and the impossibility of doing so on the ocean bed is a grave disadvantage of the scheme. Mud serves several purposes in drilling. It becomes plastered on the side of the hole and prevents material falling from it; if the mud is of high density it prevents water or oil in the rocks from being forced violently upwards and out of the hole; and it also prevents the rock chippings from falling back on the bit if circulation of the mud is stopped.

It seems imprudent to adopt a method that prohibits the circulation of mud, and there would be a great advantage in installing a pipe between the sea floor and the ship, down which the drill pipe could run, and up which the mud could flow back. It would be necessary to support this pipe with submerged floats so that it could be abandoned by the ship in bad weather or other emergency and later reconnected. If such a pipe is to be transported and installed, a ship considerably larger than *CUSS I* will be required. During the preliminary trials a prototype of the bottom support for such a pipe was installed in relatively shallow water off San Diego, but so far no tests have been made in deep water. There are considerable difficulties in handling, assembling, and maintaining so cumbrous an object as a pipe perhaps 30 cm in diameter and 4000 m long. It seems likely that further trials will be necessary before drilling of the Moho is attempted. A turbo drill with a diamond bit appears to have great advantages for a very deep hole, and will probably be used.

No site for the first Mohole has yet been selected. It should be in a 'typical' ocean floor, whatever that may mean. The Moho should be as near the surface of the ocean as possible and the heat flow should not be too high, though it is probably not necessary to seek a place where it is specially low. It is not clear whether, for a given depth of Moho below the surface of the ocean, very deep water is an advantage or a disadvantage. The weather at the site should be good for a large proportion of days in all seasons, and the site should not be too far from sources of supply of equipment or from a congenial port in which the crew can spend their periods of leave.

Considerations of weather exclude the Atlantic and the Pacific north of about 40° and also the areas of the trade winds and the hurricane belt. The heat flow off the southern Californian and Mexican Pacific coasts is higher than is desirable, but the good weather and the logistic advantages make a site somewhere in this area a likely choice.

The cost is hard to estimate, but it is probable that, before the first hole is finished, about fifteen million dollars will have been spent; the cost is unlikely to exceed twice that sum.

THE RESULTS TO BE EXPECTED

One or more holes penetrating below the Moho would resolve many uncertainties. They would show whether the Moho is a boundary between two types of rock or is a place where a change in crystal structure takes place. The rate of heat generation by radioactivity in the outer part of the mantle could be measured, which would enable us to know the limits of the wide range of temperature distributions that might, for all we at present know, occur within the earth. In particular, if the material below the Moho were found to resemble the stony meteorites in chemical composition and radioactive content, a fairly detailed account could be given of the earth's thermal history.

These things are fairly specific; perhaps the main reason for drilling the Mohole is a more general one. We have no firmly based knowledge of the history of the oceans. Within the last year papers have appeared maintaining (*a*) that the Atlantic ocean widens by continually splitting down the middle, (*b*) that all the oceans widen by a general expansion of the earth, and (*c*) that the material from the ocean floor is continually swept beneath the continents by stresses due to convection currents in the mantle. These are the views of the *avant-garde* of geology, but there are still many who believe in the permanence of the oceans, in their production by the foundering of continents, and in their conversion into continents.

These problems and most of these views have been with us for many years. The only hope of discriminating between the innumerable possibilities is to drill in the ocean floor and find what is in fact there and what is really happening. The Mohole is one of the greatest of scientific endeavours. Its achievement should lead to an outstanding advance in our knowledge of the interior of the earth.

ACKNOWLEDGMENTS

AMSOC and Dr Bascom allowed me to visit the trial drilling off Guadalupe Island last April. The ordered chaos that I saw there deeply impressed me, not only as a scientific project, but as a co-operative venture between very different kinds of people, sailors, drillers, oceanographers, and engineers among others. Without this visit I could not have written this article. I am also indebted to Dr Bascom's recent book [1] for much information.

REFERENCES

[1] BASCOM, W. 'A hole in the bottom of the sea'. Doubleday, New York. 1961.

[2] FERMOR, L. L. *Spec. Publ. Indian Ass. Cult. Sci.*, No. 6, 1938.

[3] HEEZEN, B. C., THARP, M., and EWING, M. *Spec. Pap. geol. Soc. Amer.*, No. 65, 1959.

[4] LOVERING, J. F. *Trans. Amer. geophys. Un.*, 39, 947, 1958.

[5] MOHOROVIČIĆ, A. *J. met. Obs. Zagreb*, 9, Pt. 4, Sect. 1, 3, 1909.

[6] WOOLLARD, G. P. *J. geophys. Res.*, 64, 1521, 1959.

Light quanta and vision

M. H. PIRENNE

The sensitivity of the eye is so great that it can be stimulated by a very few quanta of light. At luminances near the threshold of vision the quantum properties of light have important consequences, and they are also important in the general theory of visual processes. This article discusses some of the relevant phenomena and their theoretical interpretation. It considers also the retinal activity that may occur in total darkness and draws attention to the role of the nervous responses that occur when light is cut off from the retina.

Considered as a detector of electromagnetic radiation, the human retina has a very limited wavelength range of sensitivity, namely the visible range of the spectrum. Within this range, the retinal sensitivity[1] depends very much upon wavelength (figure 2). The sensitivity is highest when a wavelength in the neighbourhood of 500 mμ is used under such conditions that vision is mediated by the rod system found in the periphery of the retina. The numbers of quanta acting directly upon the retinal rod receptors and stimulating them are then often so small that the quantum properties of radiation have to be explicitly taken into account in the theory of visual processes, particularly when considering the purely physical fluctuations that occur in the light stimulus itself.

All the available evidence indicates that the decrease in the sensitivity of the rods for wavelengths shorter or longer than about 500 mμ is due to the fact that the photosensitive substance contained in them (rhodopsin or visual purple) absorbs proportionately less light at these wavelengths. Accordingly, the numbers of quanta actually caught by, and able to act upon, the rods remain very small when monochromatic light of wavelengths other than 500 mμ are used, and when light with a broad spectral distribution, such as white light, is used. Thus, when investigating the smallest quantity of light that will just suffice for seeing with rod vision, it is found that about a million times more quanta incident on the retina are required for light of 720 mμ than for light of 500 mμ; yet it is most probable that the number of quanta absorbed by the rods is essentially the same in both cases.

Experiments show that the quantity of light energy just sufficient for a visual response is smallest in the case of a brief flash of light of about 507 mμ observed by peripheral vision, the eye having previously been fully adapted to the dark. The energy content of one such flash is about 100 quanta; this is measured at the cornea and represents the mean amount of energy entering the eye. It is to be noted that the quantum is used here merely as a convenient unit of energy, since the quantum properties of light come into play only when light is absorbed, quanta being impossible to localize in a flux of radiant energy. Mean energy values will sometimes be expressed as fractional numbers of quanta even though, of course, the number of quanta absorbed is always an integral number. As will be shown later, the mean number of quanta absorbed by the rods in these experiments is about ten; it may possibly be less than ten, but always exceeds one. These quanta are distributed among a certain number of rods (a few hundred rods for a test field subtending an angle of 10′ of arc at the eye (figure 3)). It is impossible, if only because of the diffraction that occurs at the pupil, to confine the stimulus to one rod only. The minuteness of the quantities of energy involved in these experiments is illustrated by the fact that the mechanical energy required to lift a pea one inch would, if converted into light, be sufficient to show a faint light signal to all the men who have ever lived.

The ability of the eye to respond to the absorption of very small numbers of quanta by the retinal receptors is of practical importance in everyday life. For instance, if we wake up in the middle of a very dark night and look at the window of our bedroom, we can distinguish features of the window against the lighter patch formed by the sky. If, after a minute or two, we again close our eyes, our retinas during the whole operation will not have received enough light to enable each of the receptor cells involved to absorb even one single quantum of light. In a more general way, we may see quite well enough on dark nights to find our way about, even though at the prevailing level of illumination the majority of our retinal receptor cells remain at any instant wholly unstimulated, the sensitive substance they contain absorbing no light whatsoever. Only at illumination levels corresponding to full moonlight is the intensity enough to make each rod absorb several quanta per second.

---

[1] Spectral sensitivity, as used here is the reciprocal of the threshold quantity of light, the latter being defined as $W_\lambda \tau_\lambda / h\nu$, where W_λ is the threshold energy measurement at the cornea, τ_λ is the transmission factor of the eye media, and $h\nu$ is the quantum energy. This is the sensitivity at the retinal level on a quantum basis. The sensitivity measured at the cornea of the eye on an energy basis is $1/W_\lambda$; for the rods it is maximum at 507 mμ [31, 32] instead of near 500 mμ.

THE ROD AND THE CONE SYSTEMS IN HUMAN VISION

The retina contains two types of receptor cells, the rods and the cones (figures 3 and 4). It is the rod system which is mainly active in night vision and in all conditions when the eye is responding to low illuminations to which it has become fully adapted. After stimulation by bright daylight, it takes at least half an hour for the eye to become fully dark-adapted.

Some of the properties of the rods have already been described. The photosensitive pigment, rhodopsin, is contained in the terminal segments of the rod cells (*b* in figure 4). It is pink in colour, because it absorbs most strongly in the blue-green region and transmits the red end of the spectrum almost freely; rod sensitivity shown in curve 1 of figure 2 parallels light absorption by rhodopsin in the range where the latter can be measured [33]. In the case of dark-adapted retinas of man or of such animals as the frog, in the fresh state, this pink colour can readily be seen under the microscope under weak illumination. Under strong illumination, the colour fades rapidly as the rhodopsin is bleached by the light. It is possible to prepare aqueous extracts of rhodopsin from retinas, and its properties have been extensively studied *in vitro*.

Only quanta absorbed by the photosensitive substance can affect it; needless to say, fractions of quanta cannot be absorbed. The latter fundamental fact often simplifies the discussion of the action of light stimuli upon the retina. For instance, when the amount of energy absorbed corresponds to a number of quanta small compared with the number of receptors involved, it follows that some of the receptors will have absorbed one or several quanta, whereas the others will have absorbed no light at all. It is not quite certain that all the quanta so absorbed are 'acting quanta' in the sense that they lead to nervous excitation of the receptor cells; but in any case, the number of acting quanta can only be equal to, or smaller than, the number of absorbed quanta. It is generally accepted that what matters in rod stimulation is the number of light quanta acting upon the rods, not the energy of the individual quanta. We shall later discuss experiments made with light of about 507 mμ, but, in terms of the number of quanta acting upon the rods, the results must also be valid for other lights, for instance the light from the night sky.

The cone system of the eye is active in normal daylight, giving us then sharp and coloured vision. On account of the fixation reflex, we use direct or foveal vision to look at any point of an object which attracts our attention. The very centre of the *fovea centralis* of the retina contains cones only (figure 1). It becomes blind at low illumination levels at which the rod system, present only in the periphery of the retina, is still able to function. This is why we must—in spite of the fixation reflex—look 'out of the corner of our eye' in order to see best at night. The astronomer Arago said epigrammatically that in order to see the faintest luminous objects one must not look at them. This lower sensitivity of the foveal cone system is due in part to a smaller capacity for summating light stimuli in space and in time; this in turn gives the fovea greater accuracy of vision than the periphery. Even when small and brief test-stimuli are used, however, the smallest number of quanta required for seeing, measured at the pupil, is found to be markedly greater for the fovea than for the periphery. The foveal cone system, considered as a whole, has its maximum sensitivity near 555 mμ. It contains several cone-mechanisms having different spectral sensitivities [3]. The individual cones themselves must vary in their properties, and most probably they contain several different sensitive pigments [4]. Much less is known about the cone pigments, particularly with regard to the amount of light they absorb *in situ*, than about rhodopsin [5]. Consequently, the study of quantum effects is more difficult and less advanced for the cone than for the rod system, and in this article we shall concentrate our attention on the latter.

Whereas there are only cones in the centre of the fovea, all other retinal regions contain both cones and rods; in the periphery, however, the rods greatly outnumber the cones. Except in the case of long wavelengths, at the red end of the spectrum, vision with the dark-adapted periphery

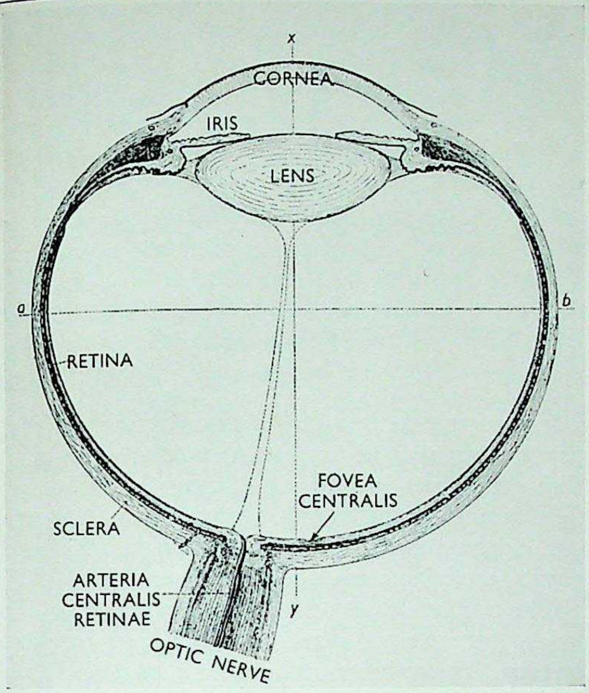


FIGURE 1 – *Section of an adult human eye. The* fovea centralis *is a depression of the retina, only a fraction of a millimetre in diameter, situated near the optic axis* xy *of the eye. Its centre contains cones only. The rest of the retina contains both rods and cones.* (×3) (*From E. A. Schäfer, 'Quain's Elements of Anatomy', Vol. III, Pt. III. London,* 1894.)

at the lowest intensities is mediated by the rod system alone. This has been demonstrated in experiments using the 'directional sensitivity of the retina' (Stiles-Crawford effect [5–7]) which is almost exclusively a property of the cones, not of the rods (figure 5). Roughly speaking, the cones are much less effectively stimulated by rays falling on to the retina obliquely instead of normally, whereas rod stimulation is, as a rule, almost independent of the direction of the incident light, within the range of directions that can be obtained using the natural pupil. Deep-red light often stimulates the dark-adapted peripheral cones as well as the rods, but by the use of a narrow beam of red light of suitable intensity, entering through the edge of the pupil, stimulation can be restricted to the rods. Suppose that in figure 5 the 'red' rays X and Y, of equal intensity, fall on the dark-adapted retina at the same spot Q, containing both rods and cones. X and Y stimulate the rods almost equally. At fairly high intensities, the ray Y, because it falls normally on the retina, stimulates the cones more than X, which falls obliquely. Now the intensity can be reduced so that, while Y still stimulates both cones and rods, X stimulates rods only. If a pencil of red light is

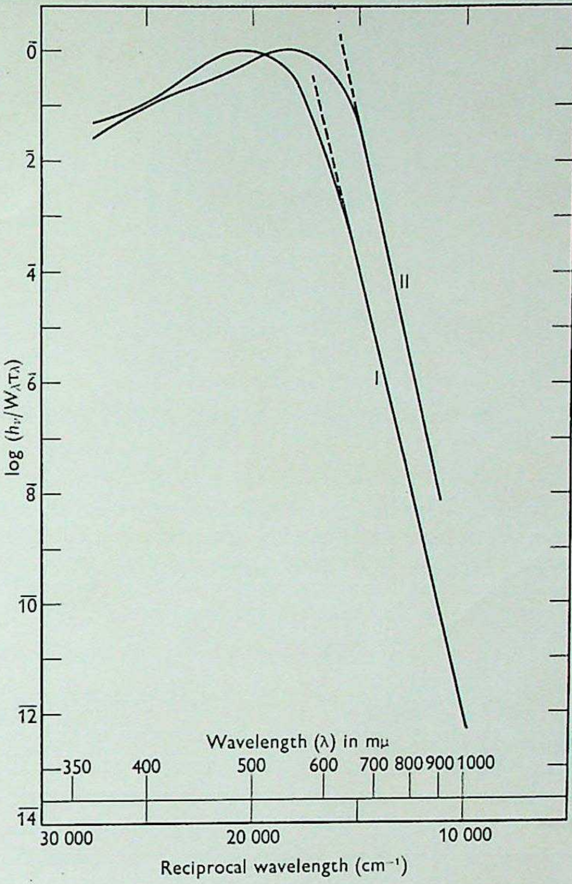


FIGURE 2 – *Relative spectral sensitivity of the human rod and cone systems. Curve I, parafoveal rod vision. Curve II, foveal cone vision. In the figure the maximum values of the sensitivity curves for the rods and the cones have been arbitrarily set equal; their actual separation on the vertical scale varies with the experimental conditions.* (*After W. S. Stiles,* [1].)

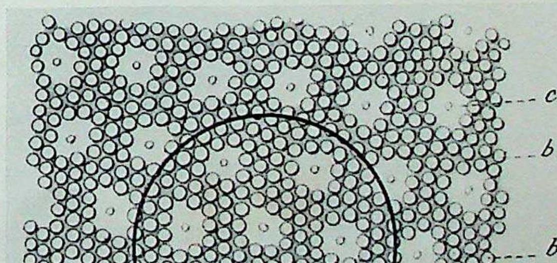


FIGURE 3 – *The mosaic of rods* (b) *and cones* (c) *in the periphery of the human retina* [2]. (×750) (*From M. Schultze,* 1866.) *The large circle represents the geometrical image,* 48·5 *microns in diameter, of a test field subtending a visual angle of* 10′ *of arc.*

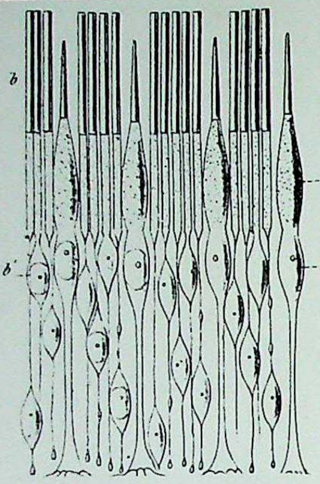


FIGURE 4 – *Section through the layer of rods* (b) *and cones* (c) *in the periphery of the human retina. The light from the pupil passes through the retina and finally reaches the rhodopsin contained in the terminal segments of the rods* b, *drawn darker in the figure.* (×500) (*From M. Schultze* [2].)

made to traverse the diameter of the expanded eye pupil, the 'red' source is first seen as almost colourless, whitish or greyish, then as red, and finally again as colourless. That is, for peripheral pupil-entry, long-wave light gives the same almost colourless sensation as given by light from the other parts of the spectrum for both peripheral and central entry. Again, the relative spectral sensitivity is found to be the same for central and peripheral pupil-entry throughout the whole spectrum except for the red region [8]. The red colour often seen with the dark-adapted periphery in the case of long-wave light stimuli, of sufficient intensity, must therefore be ascribed to the action of the cones. When the rod system is stimulated in isolation it is incapable of discriminating between different wavelengths, a finding which, among other things, gives support to the view that, within the visible range, all acting quanta cause nervous stimulation of the same kind in the rod cells, independently of the energy of each quantum. All this is related to the well-known observation that on a dark night it is impossible to distinguish the colours of objects. Red objects then look neither red nor greyish, but black, because they do not reflect enough light to stimulate either the cones or the rods.

SENSITIVITY OF ROD VISION

One of the simplest illustrations of the sensitivity of the rod system is provided by the absolute threshold of the eye for a large illuminated test-field emitting a band of wavelengths near the 507 mμ region and presented in long exposures. The conditions in one set of experiments were: field diameter subtending an angle of 47° at the eye; exposures each lasting about 15 seconds; and binocular vision with free eye-movements [9].

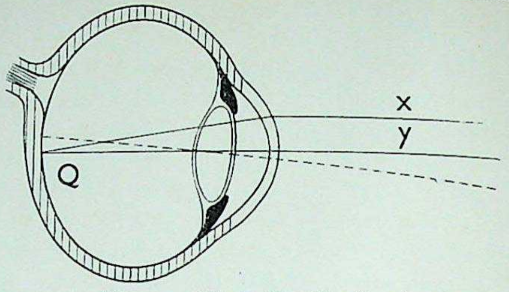


FIGURE 5 – *Diagram illustrating the directional sensitivity of the retina (Stiles-Crawford effect).* (*From W. S. Stiles,* [7].)

The absolute threshold is, broadly speaking, the lowest intensity at which the test-field can just be seen. This value must be defined on a statistical basis, however, because, for any given intensity-setting of the apparatus near the limit of visibility, the field is sometimes seen, sometimes not seen. There is no sharply defined setting below which the light is never seen and above which it is always seen. There is instead an intensity range where seeing is uncertain and unpredictable. Such a range of uncertain seeing is encountered in all experiments on visual thresholds. It must be explained by a combination of biological variation within the organism and purely physical fluctuations in the light stimulus itself. In the case of a small brief flash presented in the periphery it has been possible to show that the quantum fluctuations, which are then relatively very large, because the number of quanta acting upon the retina is very small, are a major cause of the uncertainty of seeing [10–14].

Absolute threshold is, by definition taken as the intensity at which the presence of the large illuminated field can be detected by the eye in 50 per cent of the trials, each trial lasting 15 seconds. There are variations from subject to subject in the threshold-value so defined, and even from day to day in the same subject. Similar variations are observed in the case of the statistically defined threshold-values corresponding to other experimental conditions, for instance, a small brief flash.

The average threshold-value so obtained for young normal subjects leads to the conclusion that, at the limit of visibility for our large field, each rod absorbs on the average only one quantum every hundred minutes. The energy of the quantum for $\lambda = 507$ mμ is $3{\cdot}92 \times 10^{-12}$ erg.

This conclusion rests, firstly, on a straightforward measurement, that of the mean value of the absolute energy of the light flux entering the eye through the pupil at the threshold value. This is made by determining the energy flux from the source at a higher intensity and cutting it down in the threshold measurements by the use of filters of known transmission properties. The second main basis of the above conclusion is an estimate of the fraction k of the light of 507 mμ wavelength that, having entered the eye, eventually becomes absorbed in the rhodopsin contained in the living human rod-cells. The value of k is derived (*a*) from a determination of the light losses occuring in the optical media of the human eye, from the cornea to the retina, and (*b*) from a determination of the fraction of light absorbed by the rhodopsin in the living rods. Both determinations suffer from some uncertainties, and, of course, the values probably vary to some extent from eye to eye. The fraction of the light of 507 mμ wavelength which reaches the retina through the eye media is 0·50 according to E. Ludvigh and E. F. McCarthy [15]. The fraction of this light which is then absorbed by the rhodopsin in the living rods has been estimated by W. A. H. Rushton [16] to be 0·20, on the basis of his delicate method of reflection densitometry. This is essentially a method by which the intensity of the light reflected from the back of the retina is measured before and after completely bleaching the rhodopsin *in situ* with a strong light. The difference therefore corresponds to the light absorption in a double passage through the layer of rhodopsin contained in the dark-adapted rods. On the above assumptions $k = 0{\cdot}50 \times 0{\cdot}20 = 0{\cdot}10$.

The exact value of the fraction of light absorbed by the rods is of relatively minor importance in the present context. Even on the extreme, and impossible, assumption that 100 per cent of the light entering the eye is absorbed by the visual substance contained in the retinal rods, the absolute threshold for the large field considered would still have the remarkably low mean value of one quantum absorbed per rod every ten minutes.

The value $k = 0{\cdot}10$, which should be considered as a reasonable estimate rather than as an accurate measurement, will be taken here as a basis for further discussion. It leads to a mean of $1{\cdot}73 \times 10^{-4}$ quantum absorbed per rod per second at the absolute threshold for our large field and long exposures. On the average, therefore, one quantum is absorbed every second per 5800 rods. The retinal image of a circular test-field subtending 47° at the eye has an area of about $1{\cdot}5$ cm^2. When this image falls on the retinal periphery, it covers about 750 000 cones and 20 million rods. This large number of rods absorbs an average of only 3500 quanta per second under the conditions just described.

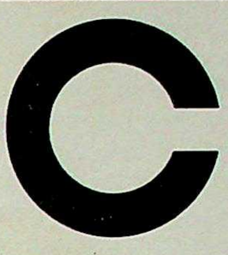


FIGURE 6 – *The Landolt ring, a test object for the study of visual acuity. The width of the gap in the ring and the thickness of the black ring are equal to one-fifth of the total diameter of the ring.*

ROD VISION AT HIGHER LUMINANCES

At the absolute threshold, the luminous field, in those trials in which it is seen at all, has an extremely vague, diffuse appearance. But if the

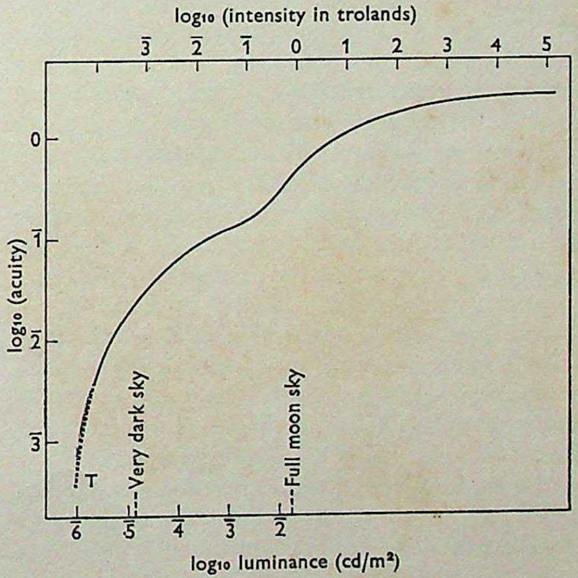


FIGURE 7 – *The general variation of visual acuity with light intensity for a black Landolt ring. Visual acuity is defined as the reciprocal of the angle, in minutes of arc, subtended at the eye by the gap in the ring. One troland is the retinal illumination produced by a surface having a luminance of* 1 *candela per square metre when the pupil area is* 1 *square millimetre. The external field luminances for the lower intensities are given on the lower abscissa scale. The point* T *indicates the average value of the absolute threshold luminance for an empty area, of* 47° *diameter, for young subjects.* [*See* 9].

light intensity is raised to three times the threshold value, the subject sees the field clearly, and can distinguish the fingers of his hand in front of it. If a black Landolt ring (figure 6), the gap of which subtends an angle of one degree at the eye, is placed in the middle of the field at ten times the absolute threshold intensity, a trained subject can detect the position of the gap in 50 per cent of the trials. Thus form-vision is here already acquiring a measurable degree of efficiency, even though the mean light-absorption per rod has only been raised to about 0·002 quantum per second—that is, one quantum every ten minutes or so. This level of intensity is a little lower than that which corresponds, in terms of quanta absorbed by the rods, to the darkest night skies [9] (figure 7).

Because the illumination values quoted above are so low in terms of acting quanta, the quantum nature of light retains its importance at higher intensity levels, where vision becomes relatively efficient. The luminance of the sky at full moon is roughly 20 000 times that corresponding to the absolute threshold of the eye. Here the position of the gap can be detected in a black Landolt ring of much smaller angular dimensions, namely when the gap subtends 3 minutes of arc at the eye. This is a fairly good degree of resolution, since the maximum resolution that the eye can achieve in bright light for the same type of object is only seven times greater, corresponding to a gap of about 0·4 minute of arc. (This maximum resolution is achieved with the central foveal cones, and corresponds roughly to the angular separation of these cones [34].) In the case of the moonlit sky, each rod on the average absorbs about 4 quanta per second. In the primary photochemical reaction, each of these quanta acts upon a single molecule, or rather upon a single chromophoric group, of rhodopsin. The number of such chromophoric groups contained in one human rod is of the order of 10^8. The quantum yield of rhodopsin bleaching is not higher than unity: that is, the ratio of the number of rhodopsin molecules that become bleached as a result of a quantum absorption to the number of molecules that have absorbed a quantum, may be unity, or possibly one-half, but not greater than unity. There is no chain reaction whereby one quantum absorbed might lead to the bleaching of several molecules; this conclusion is mainly derived from measurements of quantum yield made on solutions of rhodopsin and on rhodopsin, *in situ*, in animal retinas. Thus, as E. Baumgardt has emphasized [17], the rate of bleaching of rhodopsin in the rods remains very small over the whole range of intensity levels corresponding to night vision. At the luminance of the moonlit sky, it reaches only 4×10^{-8} per second, or 0·015 per cent per hour. Clearly the action of light upon the rods is not an ordinary photochemical reaction involving large numbers of molecules; the absolute number of rhodopsin molecules acted upon by light in each rod in night vision is small. An amplification mechanism, the nature of which is as yet unknown, must be at work in the rod receptor cell. At moonlight levels, parafoveal cones probably co-operate with rods. At higher luminances, the foveal centre becomes the most accurate part of the retina, and the cones take over.

When rod vision is effectively used, it always relies on small numbers of quanta per rod, and the total number of rhodopsin molecules contained in each rod is barely decreased by bleaching, even at the higher level of the illumination range that normally corresponds to rod vision. Towards the lower end of this range, as we have seen, the fraction of the rods actually stimulated by light per unit time is very small. The appropriate time unit is probably not less than 0·1 second; this is the approximate value of the physiological action time derived from experiments on temporal summation in absolute threshold experiments. For exposures shorter than, or equal to, this physiological action time, the mean total amount of light delivered to the eye is the determining factor. For longer exposures, the total amount of light required for threshold response increases. Assuming that the effect of one quantum on one rod cell also lasts 0·1 second, it may be said that in the case of extended fields—presented in long exposures, say of the order of one minute, at low luminances—the rods receiving the 'retinal image' of the field are for most of the time completely unstimulated. If, on the average, one rod absorbs one quantum every ten minutes or so, corresponding to a very dark night sky, it will be in a state of excitation for only about 0·1 second in every ten minutes. Certain experiments, however, suggest that light absorption by the retina may, even at intensities near the absolute threshold, lead to certain types of physiological reactions lasting as long as 5–10 seconds [18]. Rod stimulation by one quantum might last as long as this, but even so it would only occur during a fraction of the time at low intensities. As will be discussed later, experiments made with long exposures and with the eye free to move in the ordinary way probably bring into play complex physiological mechanisms. For this

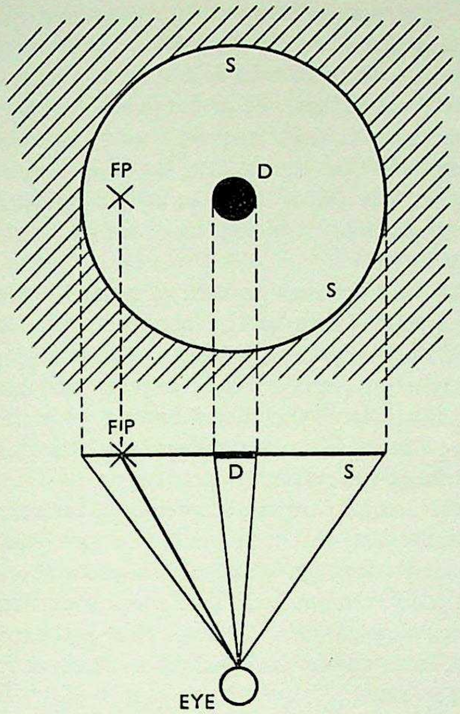


FIGURE 8 – *Arrangement for experiment by the fixation-and-flash method.*

reason, experiments made with brief flashes, shorter than 0·1 second, are often easier to interpret than experiments made with long exposures, but they no longer correspond exactly to the conditions of ordinary vision.

ACUITY OF ROD VISION, AND ITS DEPENDENCE UPON THE LEVEL OF ILLUMINATION

The familiar fact that visual acuity—that is the resolving power of the eye, in terms of form vision, or of the ability to see details—becomes higher when the light intensity increases, for such test objects as the black Landolt ring, has long remained without any adequate explanation. Some might be tempted to dismiss this problem by saying that it is obvious that we could not see well when there is not enough light; on the basis of the quantum theory, there is in fact a great deal of truth in this apparently tautological argument. For at low intensities the smallness of the number of quanta acting upon the receptors must set a natural lower limit to the size of the smallest black object, or detail of an object, that can be seen at a given intensity level. The actual limit does not seem to be very much higher than this limit, set by the properties of light itself.

In some experiments on rod visual acuity, made

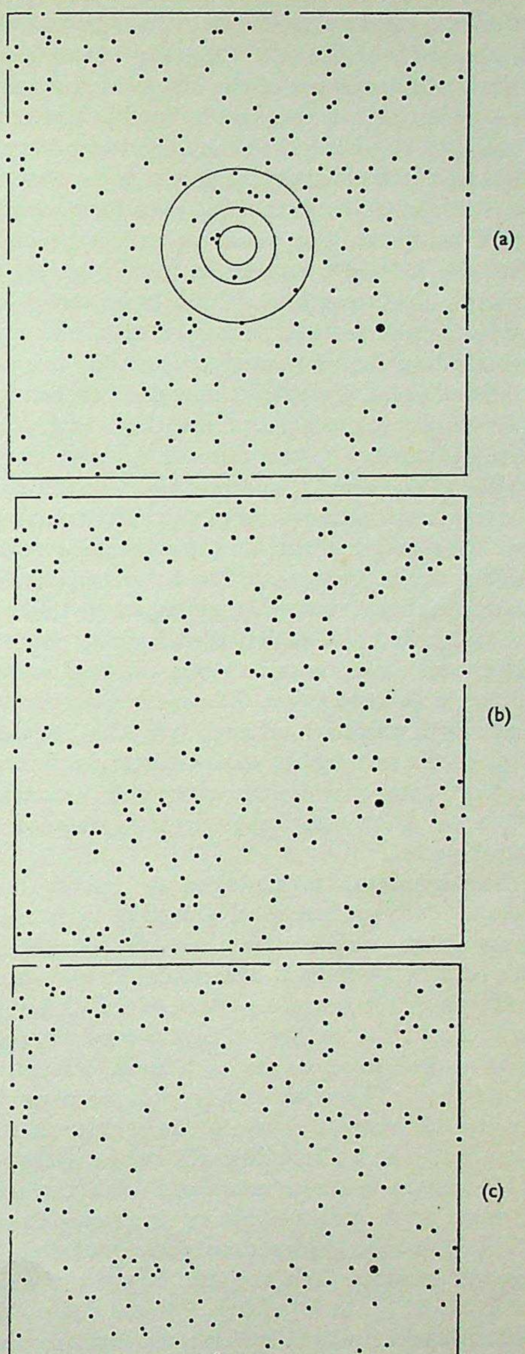


FIGURE 9 – (*a*) *Calculated random distribution of* 250 *points in a region divided into* 10 000 *small equal areas. (The large black spot indicates two points that have fallen into the same area.) The large concentric circles in the middle of the diagram are meant to represent the extent of the retinal images of three black disks of increasing diameters.* (*b*) *and* (*c*) *The same distribution of quanta as would be effected by the presence of the dark images of the two larger disks.*

with a black disk presented in the middle of a wide illuminated field, the task of the subject was merely to detect the presence of the disk, which is one of the simplest objects that can be used to test acuity of vision. We have discussed the more complicated case of the Landolt ring, partly because it is a typical test-object of the kind used for measuring visual acuity in the prescription of spectacles. With the help of a fixation point (figure 8), the luminous field, together with the black test-object, was presented always in flashes of a few milliseconds duration. The subject fixed his gaze on a small red point of light FP that was just comfortably visible in the dark. Since the image of the fixation point FP was always formed in the middle of the fovea, the retina was fixed in space—apart from inaccuracies of fixation, amounting to a few minutes of arc, that are not important in the present experiments. The subject operated a shutter and reported whether or not he had seen the black disk D—which for checking purposes was in some trials removed from the field without the subject's knowledge. These experiments were made with the dark-adapted eye, the eye being allowed to recover its dark-adaptation between flashes. Consequently, the experiment as a whole refers to a constant physiological state of the visual system.

The apparatus is arranged so that the wide field, S, surrounding the black disk, D, is evenly illuminated, yet the quanta actually absorbed by the rods in each flash are randomly distributed and form in the retina a pattern of the kind shown in figure 9. The random retinal pattern of quanta is of course different in each flash, though for convenience the same pattern has been used in all the diagrams of figure 9. It is clear that the dark image of a black disk will not be detected if its diameter is small compared with the mean distance separating neighbouring absorbed quanta. In this case, the random patterns of absorbed quanta formed when the disc is present will in practice be indistinguishable from the random patterns which would be formed when there is no black disk in the middle of the field. For each flash there are, between the rods which have absorbed a quantum (or sometimes more than one quantum), retinal areas completely unstimulated by light. The presence or absence of a dark 'shadow' on these areas cannot be detected in the experiment.

Theoretically, detection of the black disc should become possible when it produces on the retina a shadow larger than the unstimulated regions occuring naturally between the absorbed quanta. The situation may be illustrated by comparison with a dry road on which a few scattered drops of rain have fallen. If a large object, such as a motor car, is then moved away, its previous position will be detectable by the absence of raindrops, but this will not be so if a small object, such as a pebble, is picked up from the surface of the road.

Thus the quantum pattern of retinal excitation sets a natural limit to the accuracy of vision as studied in such experiments. And it must be expected that smaller black objects will become detectable when the flash intensity is higher, because the mean number of quanta absorbed per unit area of the retina is then larger.

Experimental results do agree qualitatively with this prediction, but it is generally accepted that one acting quantum is not sufficient for the eye to detect a light stimulus. The theoretical limit to the accuracy of vision, which has just been discussed, is an upper limit, which we cannot in fact expect to reach.

If one acting quantum were sufficient for vision, the light stimulus would be seen when at least one quantum had been absorbed by, and had acted upon, one rod anywhere in the retina. Statistics show that the probability of this occurring depends solely upon the total mean number of quanta absorbed by the rods and is independent of the retinal area or areas involved and of the distribution of the stimulus in time. Accordingly, if one quantum were enough for seeing, the result of absolute threshold experiments (detection of a light stimulus presented on a zero-intensity background) would be the same as if there were complete 'physiological summation' (see later) by the visual system in space and in time; in fact summation is found to occur only over a limited area and a limited time. The mean total amount of light required for seeing, as defined by the setting of the apparatus and the conditions of the experiment, is not independent of area and time. Again, it is greater when two eyes are used instead of one [19], whereas for 'one-quantum vision' it would be expected to be about the same for one eye as for two eyes, even when the two eyes belong to two different human subjects responding independently of one another [13].

We must conclude that several acting quanta are required for vision, and that these quanta must be absorbed within a limited retinal area. They must also be absorbed within a period of less than about 0·1 second—a condition which is fulfilled when brief flashes are used, as in the acuity experiments described above.

Spatial summation in the retina can be studied by using as a test object a pair of light spots placed at various distances apart. If these spots are separated by an angular distance roughly 20′ of arc or less in the periphery, their effects are to some extent additive and there is at least partial physiological summation. When the two spots are sufficiently near to one another, the threshold is determined solely by the total amount of light delivered by the two spots; this is complete physiological summation. If the two spots are far apart, there is still a greater chance of seeing at least one of the spots when both are presented together than when they are presented singly. The same increase in the probability of seeing would occur if a single spot were presented to two independent subjects, the spot being considered to be seen when at least one of the subjects has perceived it; in this case there is only 'probability summation' and no physiological summation. At the threshold, the two eyes of a subject respond very nearly on the basis of probability summation [19].

Thus there is a limited range of physiological summation of the effects of light quanta absorbed in the dark-adapted retina. Instead of considering single quanta, or rather, single rods each having been acted upon by at least one quantum, in the pattern of stimulation we must consider functional 'visual units' each containing many rods. For each unit, these rods might simply be all those that are connected to a single optic-nerve fibre (figure 12); it is also possible that the effective convergence occurs at other levels of the nervous system, such as the brain. Each of these units will be stimulated when it has absorbed a certain number of quanta, which is generally accepted to be at least two and probably not more than about ten.

Assuming, for the sake of simplicity, that we are working with a retinal region having uniform properties over the whole of its area—an assumption which is never strictly fulfilled for retinal areas of the extent considered here—a theoretical interpretation of these experiments may be attempted by considering that the dots of the diagrams of figure 9 correspond, not to single absorbed quanta, but to independent visual units, each of which has absorbed enough quanta to become excited.

The relevant point now is how many visual units have received multiple hits of sufficient magnitude to be stimulated. As long as there are, in the retinal area receiving the image of the large field, many visual units that have received too few quanta to respond at all, the situation remains relatively simple. This is because the actual levels of stimulation of the other units, which have received the minimum or more than the minimum numbers of quanta required, are irrelevant. What matters in the present theory is whether or not units are stimulated, not their degree of stimulation.

It will be noted that at this stage we are dealing, not merely with the distribution of the acting quanta among the receptors at the extreme periphery of the nervous visual system, but with the problem of the functional organization of the human retina and of the visual system. On this problem, relatively little information, at least of a quantitative nature, is available from independent sources such as anatomical studies. The problem is a complex one; for instance, many experiments on spatial summation at threshold have been made with the use of concentric test-fields of increasing diameter, instead of two light spots of increasing separation. The results of these two types of

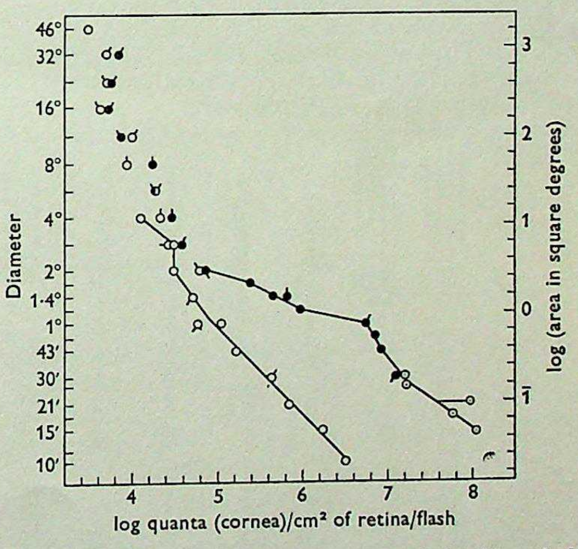


FIGURE 10 – *Results of measurements made by the method of figure* 8 *with black disks of varying diameters (black spots), and of absolute-threshold measurements made under similar conditions with luminous fields presented against a zero-intensity field (rings). All test objects centred* 20° *from the fixation point. Light flashes* 0·0026 *second in duration. Spectral band centred near* 507 *mμ. Black disks smaller than* 15′ *of arc in diameter could not be seen even at higher intensities. The straight line at* 45° *drawn through the empty circles for the smaller field diameters corresponds to complete summation, that is, to a constant total amount of light entering the eye at threshold, independently of the area of the luminous test field.* (*From P. E. Hallett,* [20].)

experiment do not agree quantitatively with regard to the spatial extent of the physiological summation, which seems to be greater in the concentric-field experiments; this fact still requires explanation. On the other hand, the quantum theory of light has helped to remove some puzzles concerning the working of the retina. Thus in the past, mere probability summation was confused with partial physiological summation, and this led to the erroneous view that physiological interaction occurred to a limited degree over the whole of the dark-adapted retina under absolute-threshold conditions.

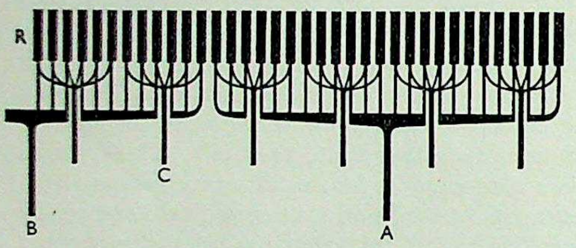


FIGURE 11 – *Schematic diagram indicating one possible type of nervous connection by which the same retinal receptors* R *would be organized into functional units of different sizes. The smaller units* C *should have a higher threshold, in terms of the illumination per unit area of the retina, than the large units* A *or* B, *since each of them could collect light only from a smaller area.* (*After Pirenne and Denton,* [21].)

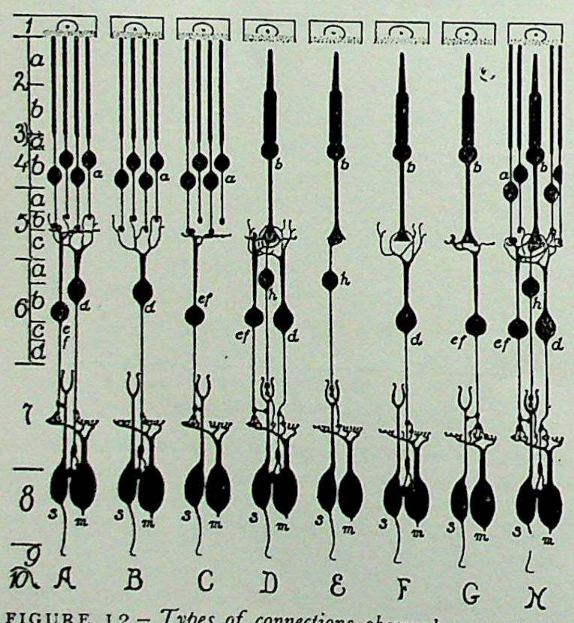


FIGURE 12 – *Types of connections observed anatomically between the rod and cone receptors and the fibres of the optic nerve. A, B, and C show how one optic nerve can be connected to several rods. The actual number of rods so connected may number tens of thousands.* (*From S. Polyak,* [22], *copyright* (1941) *of the University of Chicago Press.*)

THEORY OF THE ACUITY EXPERIMENTS

Because of quantum fluctuations, the distribution of the visual units stimulated by light will be random and vary from one flash to another; it will therefore resemble the distribution of the dots in figure 9, but here the size of the visual units is an important factor. These units probably overlap in the human retina, but there is little information on this point. It is clear that considerable complications in the quantitative aspect of the theory must be expected when the light flash becomes intense enough to cause stimulation of a large proportion of the units, and when the area of the image of the test object becomes about the same as the area of a unit, especially as the image is then likely to cover only parts of some of the units. If the visual units stimulated are few and far between, however, these complications do not arise. Under such conditions, and even at somewhat higher luminances, it is found that the threshold luminance, expressed as light intensity per unit area, is about the same for the detection of a black disk in the present experiments as for the detection of a luminous circular field having the same area—the latter arrangement corresponding, as it were, to the negative of the former, in the sense used in photography. At higher luminances, the curves for the threshold of visibility of a black disk and of a luminous circle diverge, showing that further complications begin to arise (figure 10).

The coincidence of the two curves at the lower luminances supports the theory, since it shows that the mean energy of the light that the black disk prevents from reaching the retina is the same as that of the light required to detect a luminous disk of the same area. Furthermore, in the curve relating to the black-disk experiments, a break occurs at a disk-area that is of the same order as the area within which complete summation is found to occur in the case of the luminous disk presented against a zero-intensity field (figure 10). This area probably gives the order of magnitude of the area of the visual units involved.

There are various reasons to suppose that the same retinal rods are arranged to form both large and small units, but at low luminance levels only the large units are in practice able to respond, because extremely few of the small units then receive enough quanta to be stimulated (figure 11). One would therefore expect to find a progressive recruiting of the large units as the flash luminance is increased and then, when they are all excited, a further progressive recruiting of the smaller units.

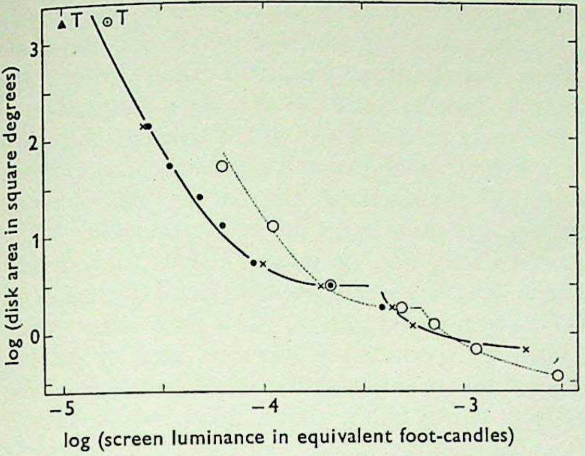


FIGURE 13 – *Results of experiments similar to those of the black disk experiments of figure* 10, *for another subject. The results are given for two retinal positions,* 20° (*solid curves*) *and* 10° (*dotted curves*) *from the fixation point. The experiments were made with white light in flashes of* 0·03 *second duration.* (*From Pirenne,* [23].)

Eventually the smallest units of the retinal periphery, consisting of cones only, or of both rods and cones, must become excited.

Figure 13 shows the results of earlier experiments performed with a different subject and under slightly different conditions. These experiments relate mainly to the limits of visibility of black disks, and cover a smaller range of disk diameters. They show that the results differ considerably at 20° and at 10° from the fovea; this was also observed in other experiments made by the subject of figure 10. Even within the periphery of the retina, the properties of the visual system vary from one position to another. This is a state of affairs observed in many types of experiments; it is certainly of practical importance so far as the use of the eyes is concerned, but it does greatly complicate the testing of visual theories.

Although the results of figures 10 and 13 are in general agreement, there are differences between them, which are probably due to differences in the experimental conditions, or perhaps to differences between the eyes of the two human subjects. Experiments of this kind, in spite of their apparent simplicity, entail a considerable effort of psychological concentration on the part of the subject. They are also time-consuming, since each experimental point is given by a statistical determination based on results corresponding to many flashes. After a couple of hours, the subject's fatigue tends to lead to irregularities in his responses. Consequently, the number of points that can be determined at one sitting is limited. Furthermore, on account of day-to-day variations in the same subject, it is not always easy to co-ordinate the results obtained at different sittings. Finally, special precautions must be taken to prevent the experimenter from unwittingly influencing the subject's results. The well-known story of the alleged discovery of the 'N-Rays' has demonstrated the effect psychological suggestion can have on careful and completely honest visual observers working under difficult conditions [24].

The above results constitute no more than one step in the study of visual acuity, and they have been discussed here because of the importance of this visual function.

SUBJECTIVE APPEARANCES

In absolute-threshold experiments made with a fairly large luminous field whose presence must be detected against a zero-intensity field, it is tempting to assume that the field is just visible when only one of the visual units covered by its retinal image is stimulated. But though this assumption is often made, there is no proof that it is true. The 'two-quanta theory' developed by H. A. van der Velden [12] and M. A. Bouman [30] is based on the assumption that the condition for seeing is that one visual unit be acted upon by at least two quanta—the number two being fixed, according to the theory. On the other hand, a more involved theory of the absolute threshold has been put forward by Baumgardt [25], according to which seeing a test field depends on more complicated 'coincidence' arrangements than the above hypothesis of physiologically independent visual units.

The subjective appearance of the field is of little immediate use in testing the above assumption. It might be thought that when only one, or two or three, units are excited within the area corresponding to a fairly large test field, the subject should see one, two, or three discrete patches of light in definite positions. But reasonings of this kind are notoriously unreliable, because, for one thing, there would be no practical or biological purpose in our 'seeing' the actual stimulation pattern in our retina. Some experiments were made in which a subject was presented (*a*) with a large circular field exposed in flashes at a fixed luminance near threshold; (*b*) with the top half only; or (*c*) with the bottom half only, of the field [26]. The experimental conditions, including the luminance of the fields, remained the same, and the actual sequence of the presentation of (*a*), (*b*), and (*c*) was unknown to the subject.

The subject then made confusions between the half-fields (*b*) or (*c*) and the whole field (*a*); he mistook, at times, a half-field for the whole field. This is in keeping with the theory that only a few, sometimes perhaps only one, of the visual units covered by the field are stimulated at threshold. Then it must be expected that the stimulation caused by the whole field will sometimes be confined, say, to the top half only, and this will explain the confusions that occurred.

It would be more straightforward to find out whether the whole field can be confused with a single small test-field covering an area about the size of the retinal area inside which physiological summation is known to take place, or with random arrays of a few such areas. But the same luminance must be used for the whole field as for the small fields. If a higher luminance is used for the small fields, the stimulation of the visual units concerned is likely to be higher in the latter, and this may introduce a significant difference between the two modes of stimulation, for it is clear that a few really bright spots of light are not confused with a large extended field. Here the difficulty arises that if only one or a few small fields are used at a luminance near the threshold luminance (50 per cent frequency of seeing) of the large field, the frequency with which the subject's threshold is reached in the case of the small fields becomes so low that in practice he hardly ever has the opportunity of comparing their appearance with that of the large field. This effect on the threshold luminance is readily explained on the basis of probability summation.

In the case of the black-disk experiments also, the luminous field is not seen as a pattern of light spots; the unstimulated regions predicted by the theory are not perceived subjectively. At luminances well above that at which the large luminous field itself can just be seen, but adjusted to a luminance below that necessary to see the black-disk, the disk of course remains entirely invisible. Yet subjectively the field may then appear quite bright and very evenly illuminated. This will be so, for example, for a black disk 1° in diameter, twice the angular diameter of the sun; judging from the subjective appearance of the field, a subject unused to the experiments may find it hard to believe that he failed to see such a conspicuous object in the experiments.

Apart from the difficulty of maintaining attention, one of the problems confronting the human subject in absolute-threshold experiments is that of avoiding confusion of the test field with the ever-changing luminous clouds, the 'phosphenes', are subjectively seen in total darkness. Some subjects are much less troubled by phosphenes than others, but darkness as a rule never looks completely dark or black to the dark-adapted eye. The visual field is filled with *Eigenlicht*, the eye's own light. No doubt such effects are due to some form of spontaneous activity in the nervous system, but their exact origin is uncertain. They may, or may not, be related to the 'dark noise' which, according to some theories [27], exists in the human visual system. This dark noise would consist of spontaneous nervous activity having characteristics indistinguishable from those of the activity caused by the action of light on the retina. It might, for instance, be due to the thermal activation of rhodopsin molecules in the rods, on the assumption that such activation causes the same effects as the activation due to actual light absorption. There is no reason to suppose that such dark noise would necessarily lead to luminous sensations in complete darkness, as these would be of no apparent biological use. But since, by definition, the dark noise would be of the same kind as the nervous stimulation caused by light, it would necessarily interfere with the detection of actual light stimuli if the noise were of sufficient magnitude. This might explain why more than one acting quantum is necessary for vision—even though the strange effects which would then occur with regard to summation might also make one-quantum vision biologically useless. At the present time there does not seem to exist any convincing proof of either the presence or absence of such dark noise at a significant intensity in the human visual system.

CONCLUSION

Most of the experiments discussed above were made with brief light-flashes. With regard to visual acuity, the processes involved in ordinary vision are probably much more complicated. Then the retina becomes adapted to the prevailing illumination, and the eye moves continually, making fixation pauses of the order of 0·1–0·2 second in the course of its jerky displacements. Under these conditions it is probable that the perception of dark objects involves 'off' responses in the visual system; that is, nervous responses that occur when the light is cut off from the retina. Such 'off' discharges of impulses have been observed in the case of electrophysiological preparations of vertebrate eyes, for instance in optic-nerve fibres of the frog [28]. They may occur in the human retina when, as a result of a displacement of the eye, the image of a dark object suddenly falls on a previously illuminated area. In this connection, it must be remembered that there is much evidence indicating that stimulation of the retina under steady conditions leads to very inefficient vision, or to complete loss of vision, after a few seconds. Accordingly, it may be surmised that, when the eye looks at a large evenly illuminated field, the central part of this field soon becomes visually ineffective, since it is only near the edges of the large field that eye movements cause changes in the illumination reaching the retina. If so, the only means of affecting the retina when the eye tries to detect a small black object in the centre of a large field would be displacements of the dark image of the object on the retina. There are reasons for believing that over the whole range of rod vision, and over a great deal of the range of cone vision, it is the cutting off of the retinal illumination by this dark image that is the determining factor for the response of the visual system at the limit of visibility of the black object [29].

The experiments discussed previously can do no more than indicate the ways in which the conception of quanta of light can be applied to the study of the visual process. The quantum theory does not solve the problems of the physiology of vision, but it allows many of them to be studied in a more concrete and exact manner, and so helps to reveal their complexity.

ACKNOWLEDGMENTS

Some of the experimental work reported in this article has been supported by the Medical Research Council, the Nuffield Foundation, and the United States Public Health Service (Grant No. B-1810 from the National Institute for Neurological Diseases and Blindness).

REFERENCES

[1] STILES, W. S. 'The physical interpretation of the spectral sensitivity curve of the eye'. In 'Transactions of the Optical Convention of the Worshipful Company of Spectacle Makers'. Pp. 97–107. Spectacle Makers' Company, London. 1948.

[2] SCHULTZE, M. *Arch. Mikr. Anat.*, **2**, 175, 1866.

[3] STILES, W. S. *Proc. roy. Soc.*, *B*, **127**, 64, 1939.

[4] RUSHTON, W. A. H. *Progr. Biophys.*, **9**, 239, 1959.

[5] STILES, W. S. *Ann. R. Coll. Surg. Engl.*, 1961. (In press.)

[6] STILES, W. S. and CRAWFORD, B. H. *Proc. roy. Soc.*, *B.*, **112**, 428, 1933.

[7] STILES, W. S. *Sci. Progr.*, **33**, 676, 1939.

[8] FLAMANT, F. and STILES, W. S. *J. Physiol.*, **107**, 187, 1948.

[9] PIRENNE, M. H., MARRIOTT, F. H. C., and O'DOHERTY, E. F. *Spec. Rep. Ser. med. Res. Coun. Lond.*, No. 294, 1957.

[10] HECHT, S., SHLAER, S., and PIRENNE, M. H. *J. gen. Physiol.*, **25**, 819, 1942.

[11] BRUMBERG, E. M., VAVILOV, S. I., and SVERDLOV, Z. M. *J. Phys. Moscow*, **7**, 1, 1943.

[12] VAN DER VELDEN, H. A. *Physica*, **11**, 179, 1944.

[13] PIRENNE, M. H. *Progr. Biophys.*, **2**, 192, 1951.

[14] PIRENNE, M. H. and MARRIOTT, F. H. C. In 'Psychology: A Study of a Science. Study 1. Conceptual and Systematic. Vol. 1. Sensory, Perceptual, and Physiological Formulations'; S. Koch (Ed.). Pp. 288–361. McGraw-Hill, New York. 1959.

[15] LUDVIGH, E. and MCCARTHY, E. F. *Arch. Ophthal. N.Y.*, **20**, 37, 1938.

[16] RUSHTON, W. A. H. *J. Physiol.*, **134**, 30, 1956.

[17] BAUMGARDT, E. 'Les théories photochimiques classiques et quantiques de la vision et l'inhibition nerveuse en vision liminaire'. Editions de la Revue d'Optique, Paris. 1950.

[18] Ref. [11], p. 56.

[19] PIRENNE, M. H. *Nature, Lond.*, **152**, 698, 1943.

[20] HALLETT, P. E. *Proc. 3rd Int. Congr. Photobiol.* (Finsen Memorial Congress, Copenhagen). (In press.)

[21] PIRENNE, M. H. and DENTON, E. J. *Nature, Lond.*, **170**, 1039, 1952.

[22] POLYAK, S. 'The Retina'. University of Chicago Press, Chicago. 1941.

[23] PIRENNE, M. H. *Proc. Camb. phil. Soc.*, **42**, 78, 1946.

[24] PIÉRON, H. In 'Mechanisms of Colour Discrimination'. P.15. Pergamon Press, London. 1960.

[25] BAUMGARDT, E. *Proc. Int. Illum. Comm. XIV* Session, Brussels, Vol. A, p.78, 1959.

[26] PIRENNE, M. H. *J. Physiol.*, **107**, 47*P*, 1948.

[27] BARLOW, H. B. *Ibid.*, **141**, 337, 1958.

[28] HARTLINE, H. K. *Amer. J. Physiol.*, **121**, 400, 1938.

[29] PIRENNE, M. H. *Ann. N.Y. Acad. Sci.*, **74**, 377, 1958.
MARRIOTT, F. H. C. and PIRENNE, M. H. *J. Physiol.*, **152**, 24*P*, 1960.

[30] BOUMAN, M. A. *Studium gen.*, **13**, 491, 1960.

[31] CRAWFORD, B. H. *Proc. Phys. Soc. Lond.*, *B*, **62**, 321, 1949.

[32] *Proc. Int. Illum. Comm.*, **3**, 32, 1951.

[33] CRESCITELLI, F. and DARTNALL, H. J. A. *Nature, Lond.*, **172**, 195, 1953.

[34] SHLAER, S. *J. gen. Physiol.*, **21**, 165, 1937.

GENERAL READING

PIRENNE, M. H. 'Vision and the Eye'. Chapman and Hall Ltd, London. 1948.

Idem. Biol. Rev., **31**, 194, 1956.

MARRIOTT, F. H. C. *Symp. zool. Soc. Lond.*, No. 3, 67, 1960.

BRINDLEY, G. S. 'Physiology of the Retina and Visual Pathway'. Edward Arnold, London. 1960.

PIRENNE, M. H. and MARRIOTT, F. H. C. 'Visual Functions in Man'. In 'The Eye', H. Davson (Editor), Academic Press, London and New York. (In press.)

National Physical Laboratory Symposium No. 8, 'Visual Problems of Colour'. 2 vols. Her Majesty's Stationery Office, London. 1958.

Sir James Hall, Bt. (1761–1832)

V. A. EYLES

Sir James Hall was a friend both of Lavoisier and of James Hutton, the geologist who first advanced the modern theory of the origin of the igneous rocks. Hall devised a number of chemical experiments to test the validity of the theory by attempting to reproduce subterranean temperatures and pressures; he was the first geologist to use such methods successfully. The experiments, and the results they achieved, are described in this article.

Sir James Hall (figure 1), who was born on 17th January 1761, was the first person successfully to apply the experimental techniques of the chemist to geological problems. The experiments were carried out between 1798 and 1804, when the use of such methods by geologists was almost unknown. They were made to test the theories put forward by Hall's friend, James Hutton, whose 'Theory of the Earth' was first announced in 1785 [1].

A short account of Hall's career will serve to show how he became involved in Hutton's problems. He was a descendant of an old Scottish family established in East Lothian. In his twelfth year he was sent to a school near London, and at the age of fifteen he succeeded to the baronetcy. Sir John Pringle, at that time President of the Royal Society, was his great-uncle, and he took young Hall to dine at the Royal Society Club, a dining club within the Royal Society, in London. This may have stimulated Hall's interest in science; but it is probable that he had already met Hutton, who was a neighbour and friend of his father. In 1777 Hall entered Christ's College, Cambridge, where he remained in residence for two years without graduating. After leaving Cambridge he studied first in France and then in Geneva. He returned to Scotland for his coming-of-age in 1782. In 1783 he went abroad again, for about three years, spent travelling in Europe. He was now actively interested in geology—during this tour he visited Vesuvius and the volcanoes in Sicily and the Lipari Islands. In Sicily, he was accompanied by a distinguished French geologist, Déodat Dolomieu. Hall made a careful examination of Vesuvius and the adjacent Monte Somma, and the sketches and notes he made at the time proved useful when he came to study Scottish geology. During 1786, he spent several months in Paris, and visited Lavoisier frequently. As a result, he became a convert to the latter's views on chemistry.

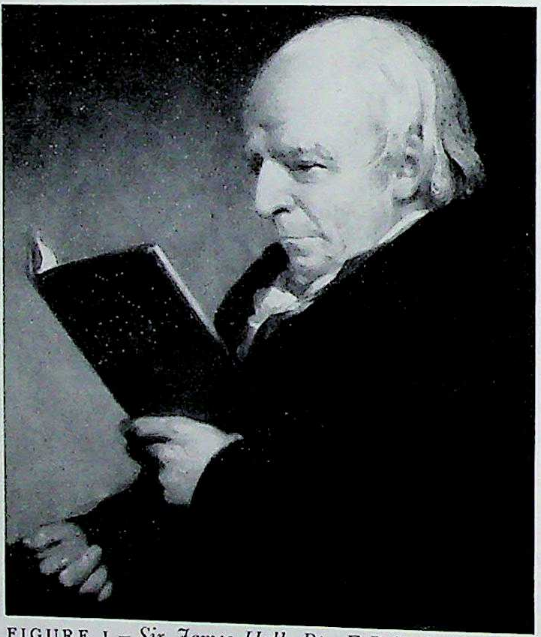


FIGURE 1 – *Sir James Hall, Bt., F.R.S., by Sir John J. Watson-Gordon, R.A.*

He returned to Edinburgh in November 1786 to marry Lady Helen Douglas. Early in 1788 he addressed the Royal Society of Edinburgh on 'Lavoisier's New Theory of Chemistry'. This discourse, occupying three meetings, roused both interest and opposition, for two subsequent meetings were devoted to an exposition of the phlogiston theory by Hutton, and by Hall's reply to Hutton. Hall's address was not published, but it was mentioned by Robert Kerr, the translator of Lavoisier's *Traité élémentaire de chimie*, in a letter written to Lavoisier on 21st January 1791. Kerr wrote that Hall had given the Royal Society of Edinburgh a very detailed account of Lavoisier's doctrine and had converted some people, including Kerr himself, to the antiphlogistic theory [2]. This letter shows that Hall was one of the earliest supporters of Lavoisier's doctrines in Britain. He had been deeply interested in chemistry since his

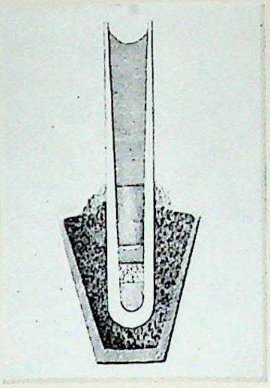


FIGURE 2 – *Porcelain tubes as used in the second stage of compression experiments.*

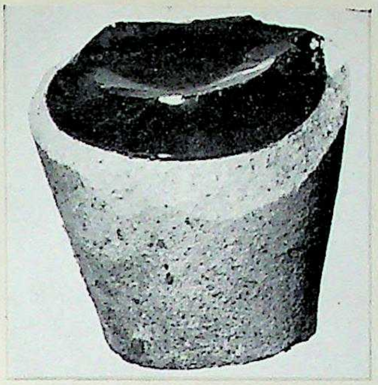


FIGURE 3 – *Original crucible, used by Hall in his 'whinstone' experiments, showing the glass formed by rapidly cooling a melt of igneous rock (height about 6 cm).*

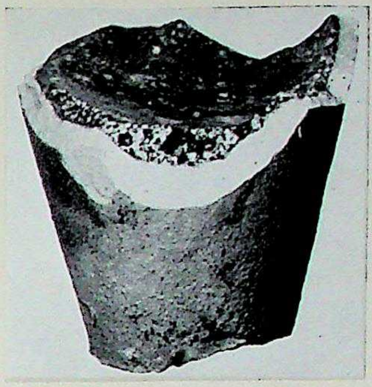


FIGURE 4 – *A similar crucible, showing the crystalline mass obtained by slow cooling.*

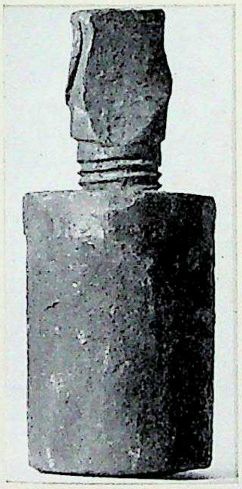


FIGURE 5 – *Drilled iron bar with plug. (7·5 cm long.)*

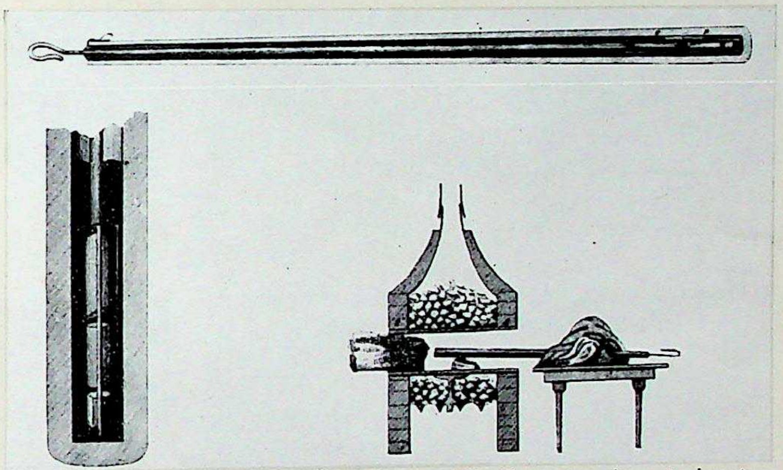


FIGURE 6 – *Earliest apparatus used by Hall in his compression experiments.*

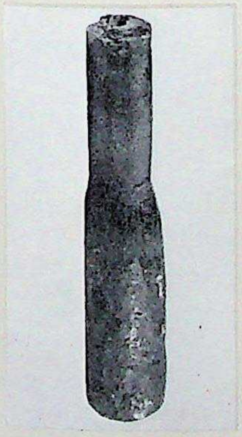


FIGURE 7 – *Drilled bar (37 cm long).*

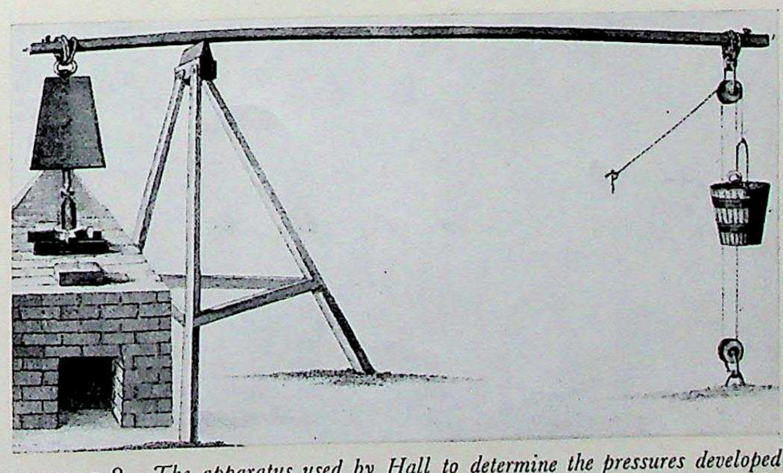


FIGURE 8 – *The apparatus used by Hall to determine the pressures developed during his experiments. A very heavy iron weight rests on the open end of a tube filled with fusible metal, and the reading is taken when the expansive forces of the gases generated start to lift the weight. Exact adjustments are made by counterpoise weights placed in the bucket at the end of the beam.*

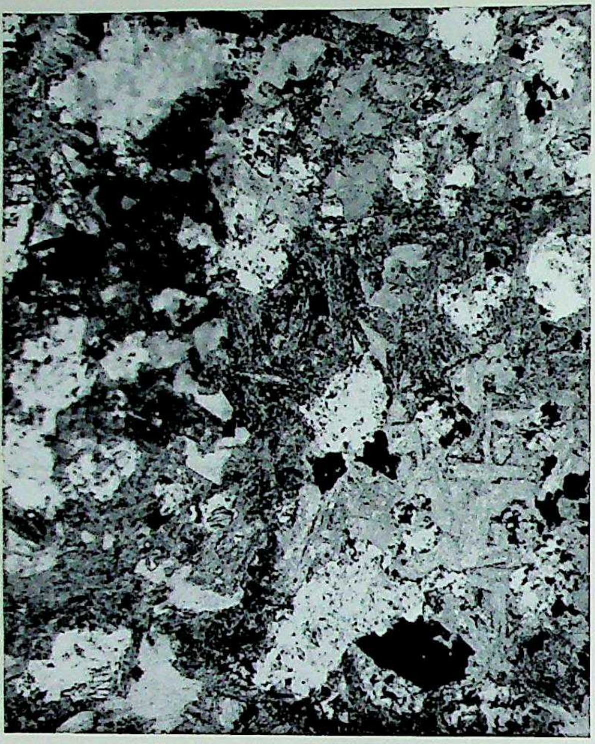


FIGURE 9 – *Thin slice of one of the 'whinstones' used by Hall, before melting, showing typical crystalline texture.*

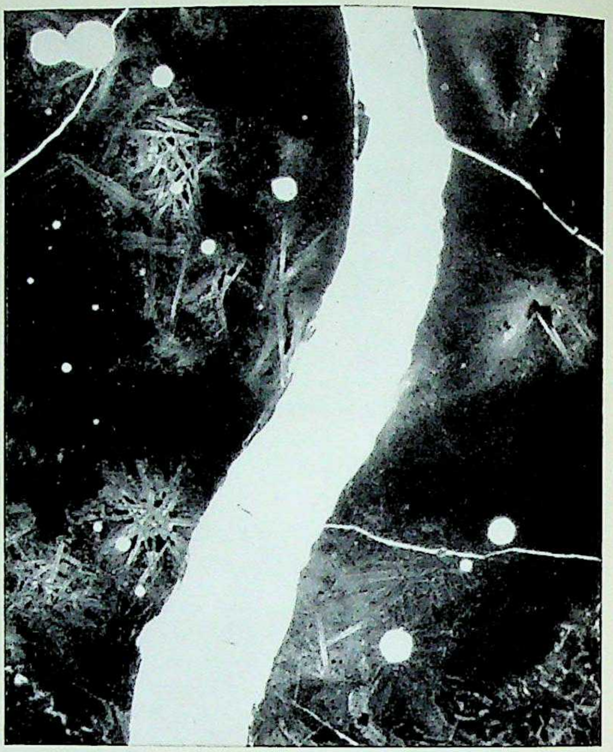


FIGURE 10 – *Thin slice of slowly cooled igneous melt, showing crystallites in a glassy base.*

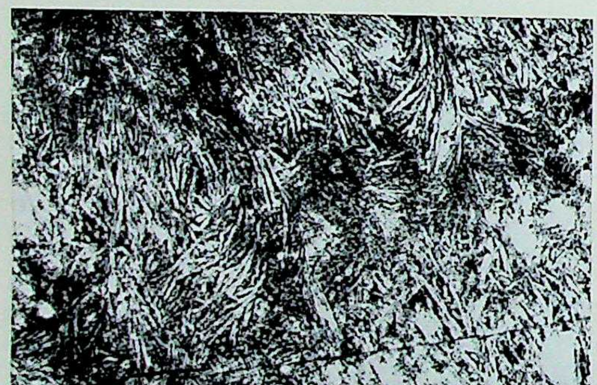


FIGURE 11 – *Another slowly cooled igneous melt. The texture is not that of a typical unmelted whinstone (see figure 9). (Figures 9–11 highly magnified.)*

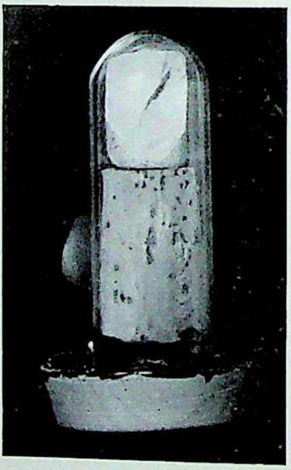


FIGURE 12 – *A specimen of 'marble' produced in one of Hall's experiments.*

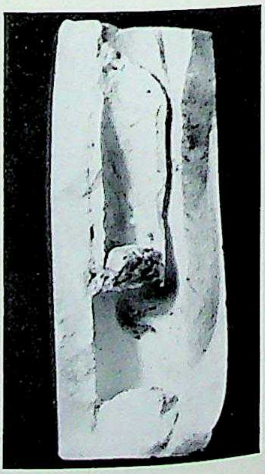


FIGURE 13 – *Porcelain tube, sectioned, after an experiment.*

student days, and had attended the lectures of Joseph Black in Edinburgh.

It was at this time, in 1788, that the first full account of Hutton's 'Theory' was published. Hall's first reaction was unfavourable. He said, later, that his first inclination was to reject the theory, and that he would have done so had he not known the author, whose lively, perspicacious conversation was strikingly contrasted to the obscurity of his published writings. It was only after three years of almost daily conversation that Hall began to accept Hutton's fundamental principles. The change in opinion was a fortunate one, for it led Hall to his classic experiments.

Some knowledge of the state of geology at that time is necessary for an appreciation of the experiments. Many geological facts had been accumulated—by observation of surface rocks, by mining operations, and by the study of minerals; what had hitherto been lacking was a master theory that would co-ordinate the known facts and explain their significance. Much was already known about the sedimentary rocks—the limestones, shales, and sandstones—but there was another class of rocks, including the crystalline granites, basalts, and porphyries, about whose origin the geologists were almost completely ignorant. These are now called 'igneous' rocks. We know today that they originated in the Earth's interior, and that they have consolidated from a molten condition. The widely held view in Hutton's time was that, like the sediments, they had been laid down in water, or had crystallized from solution. It is remarkable, considering how much was then known about the solubility of chemical compounds, that this view should have been held, for the rocks are mainly composed of virtually insoluble silicate-minerals, such as quartz and felspar. A single exception to the general view was provided by lavas, a term then restricted to rocks that had obviously flowed from active or recently extinct volcanoes; most geologists thought that this phenomenon was purely local. They believed that many basalts in Europe and elsewhere, now known to be lavas, were aqueous deposits. This view continued to be held by some prominent geologists until the early years of the nineteenth century.

The first correct and comprehensive explanation of the origin of the igneous rocks was put forward by Hutton. He supposed that in the deeper parts of the Earth's crust all rocks were fused, and that great quantities of molten rock had moved up into the upper sedimentary parts of the crust. Some poured out on the surface as lavas and some consolidated underground. The latter formed 'subterraneous lavas' (flat sheets), vertical dykes (veins), or large irregular masses, such as the granites of Scotland. This was a major advance in geological thought, and those who accepted Hutton's views on the origin of igneous rocks came to be known as 'Plutonists', in opposition to the 'Neptunists', who held the earlier views.

Hutton's theory was much criticized on particular points, and we are especially concerned here with two criticisms that carried considerable weight at the time. They arose from ignorance of the chemistry of rock melts. This is not surprising, for the subject is very complex.

The first criticism concerned the fusion of igneous rocks. It was known that some of the commoner igneous rocks, when melted in a crucible and allowed to cool, formed a glass, not a crystalline rock. The Neptunists claimed that, if the rocks that Hutton supposed once to have been molten really had consolidated from such a condition, they would now be found as glasses and not as crystalline rocks.

The second criticism arose from the fact that limestone, or carbonate of lime, when heated, decomposes to quicklime, giving off carbon dioxide. This was a fact particularly well known to Edinburgh scientists, for it had been established by Black. Hutton now claimed that rocks forming land surfaces, including limestones, had, in the past, been raised to the surface after deep burial in subterranean regions where very high temperatures prevailed. Why, the critics asked, had the limestones not been decomposed?

Hutton's answer to both objections was that the conditions existing deep in the Earth's crust were very different from those on the surface; he thought it impossible to imitate them. He appears to have feared that experimental attempts to meet criticisms might do more harm than good if the experiments were unsuccessful. He later issued a warning against those who 'judge of the great operations of the mineral kingdom from having kindled a fire and looked into the bottom of a little crucible'.

Hall was, of course, aware of Hutton's views, but it seems that he still thought it worth while attempting to test the theory experimentally. He carried out a few preliminary experiments in 1790, but discontinued them out of respect for his friend's opinion. Hutton died in 1797 and Hall resumed his experiments.

His first were his 'Experiments on Whinstone and Lava' [3]. He hoped to show that a melt of igneous rock, if cooled slowly, would produce a crystalline rock, and not a glass. He used the Scottish term 'whinstone' for rocks that, in Hutton's view, were igneous intrusions, and the term 'lava' in its then restricted sense, for rocks known to have been extruded by a volcano. It had already been demonstrated by James Keir [4] that crystallization could take place in ordinary glass if cooled slowly; and Hall had been reminded of the fact more recently by seeing, in a local glass works, the result of slow cooling of a melt of green bottle glass, which had become a stony mass. He believed, therefore, that his chances of success were good. He used eight different 'whinstones' and seven lavas, some of which he had brought back himself from Vesuvius and Sicily. Each specimen was first heated in the reverberatory furnace of an iron foundry until thoroughly molten, and then cooled rapidly to a glass. Fragments of the glass so formed were then reheated until molten, and held for some time at a lower temperature in an open fire or in a muffle furnace before being allowed to congeal. Hall soon found that by regulating the rate of cooling he could produce a stony, more or less crystalline, mass. He discovered by trial and error how long to keep his melts in the muffle furnace—he mentions twelve hours for one experiment. The preliminary conversion to glass was a precaution against the criticism that he had not heated his specimens sufficiently to destroy their crystalline texture (figures 3 and 4).

It seems clear that in all Hall's experiments the final product showed some evidence of the presence of crystals. In some it was only incipient crystallization, in others he obtained rocks that may have resembled superficially those with which he had started. The only means he had for examining them was a hand lens; the use of thin rock-sections had not then been developed. Fortunately, some of his material is now in the Geological Museum, London. This has been examined microscopically, and, in one melt, crystals of felspar, augite, olivine, and iron ore were recognized, minerals characteristic of some of the rocks melted. In others, though crystallites or centres of spherulitic crystallization had developed, a glass matrix remained. Hall does not seem to have succeeded in producing a rock with a crystalline texture comparable to that of the rocks with which he started (figures 9–11). This is not surprising, for the time taken for naturally occurring igneous rocks, even lavas, to cool is very much more than that allowed by Hall.

Hall kept a check on the temperatures reached during his experiments by means of 'thermometric pieces', or pyrometers. These were invented by Josiah Wedgwood the elder for measuring temperatures too high to be recorded by any thermometer then available. Unfortunately, it is impossible to convert the figures on Wedgwood's scale to those on standard thermometers; but in some experiments his material was raised to temperatures exceeding the melting point of silver, about 950° C.

Hall's experiments on whinstones and lavas were interesting and valuable, but his next investigation, on 'The Effects of Compression in Modifying the Action of Heat' [5], was more important. He heated carbonate of lime to high temperatures and under high pressures to demonstrate that under these conditions it would not dissociate to quicklime and carbon dioxide. He also hoped to form crystalline marble, and even to melt the carbonate of lime. He carried out, between 1798 and 1805, more than 500 experiments. It was a process of trial and error which, as his experimental skill increased, led finally to success. Hall describes the methods and the results in much detail; it is possible only to summarize his account here.

He experimented with different kinds of calcium carbonate. He used, generally, powdered chalk; but he also worked with, for example, common limestone, marble, and chemically pure carbonate of lime. Hall realized that the carbonate of lime that occurs naturally is seldom pure. The amounts he used for each experiment varied from about half a gram to five grams.

His earliest attempts were made with a container with a lid fixed with screws, but this method was soon abandoned for a much simpler one. This used iron tubes of narrow bore, sealed at both ends. These were, in fact, gun-barrels, the breech being closed by welding in an iron plug. The carbonate, wrapped in paper, was inserted into the breech, and the rest of the barrel was then filled with baked and powdered clay. The muzzle was closed with another iron plug, and the end of the barrel containing the carbonate was put into a muffle furnace, the muzzle remaining outside, where it was kept cool with wet cloths (figure 6). One of the earliest results was the explosion of the furnace, but this did not deter Hall. He constantly modified his apparatus. He was not satisfied with the results obtained using powdered clay to fill the barrel, and substituted a fusible metal that melted at the temperature of boiling water and could easily be removed after the completion of an experiment. This method was used for some time, until another difficulty arose. The fusible metal sometimes diffused through the iron barrel, from which it was extruded as fine wire. This, Hall thought, must have been the result of the pressure set up by expansion of the metal. He therefore inserted with the carbonate a small tube, open at one end, containing air, which, he thought, would act as a cushion to the expansion of the fusible metal. This method seems to have succeeded, once he had discovered, after destroying many gun barrels, the correct amount of air to use. He prevented contamination of the carbonate by placing it in small tubular containers, about 6 mm in diameter. He used glass first; but this proved too fusible, and he substituted porcelain. The containers were held in an iron frame on the end of a ramrod, so that they could easily be removed after an experiment (figure 6). He sometimes used lead instead of the fusible metal, and he also put powdered 'silex' (probably flint or quartz) into the porcelain container to prevent direct contact between carbonate and fusible metal. A major change was the substitution of porcelain tubes (about 34·5 cm long, having a bore of 6 mm and walls 5 mm thick), obtained from Wedgwood's Etruria works, for the gun-barrels previously used (figures 2 and 13). He also used silex as a filler, and left a space at the open end of the tube to be filled with melted borax to make the tube gas-tight, but this proved unsatisfactory. Hall suspected that the porcelain tubes were porous, and he used melted borax, with some success, as an internal glaze. Later, being dissatisfied with Wedgwood's tubes, he manufactured more satisfactory ones himself from a mixture of Cornish porcelain clay and 'Cornish stone' (possibly china clay). He varied his method of placing the tubes in the muffle furnace, finding it more satisfactory, when using borax as a seal, to insert them vertically (figure 2). The increased gas-tightness permitted higher pressures to be used.

By 1803 Hall had carried out 156 experiments, with variable results; in some, the loss of carbon dioxide was quite small, 2–3 per cent, or even less. At this stage he was quite justified in concluding that calcium carbonate heated under pressure was not significantly dissociated even at temperatures above 1000° C. It also seems probable that he had succeeded in converting powdered chalk into something closely resembling crystalline marble (figure 12). In one experiment, using 80 grains (about 5 grams) of powdered chalk, he produced a compacted semi-transparent mass, the interior of which had the characteristic rhomboidal cleavage of calcite. The loss in weight was only 3·5 grains, less than 5 per cent.

Hall still hoped to show that the calcium carbonate could be melted. He decided to increase both temperature and pressure, and decided that the porcelain tubes he had been using would not prove satisfactory. He tried gun-barrels again, but found them unsuitable too. So he decided to use iron bars, in which he drilled holes for the porcelain containers (figures 5, 7). He was able to obtain a supply of bars of an exceptionally high-grade iron, known as 'Old Sable' iron, which was imported from Russia. A friend, Dr Robert Kennedy, a skilled chemist, suggested he should place in the barrel, with the carbonate, small amounts of water, which, on vaporizing, would greatly increase the pressure developed. Hall adopted this suggestion, and even thought he could improve on it by using, in some experiments, ammonium carbonate, ammonium nitrate, gunpowder, or paper impregnated with nitre, but he found them no improvement on water. One result of these methods was an interaction between the carbonate and the clay container, forming one or more new minerals, which appear to have been rendered molten, or to have been redeposited after sublimation.

In 1804 Hall publicly claimed that he had succeeded in fusing the carbonate. It was suggested that this may have been the consequence of the presence of impurities, so he decided to use platinum as a container. He could obtain it only in the form of thin sheets, which he folded like a filter paper, to make a cup. He also obtained a supply of specially prepared, chemically pure, carbonate of lime.

As in his whinstone experiments, Hall kept a careful check on the temperatures reached by means of 'pyrometer pieces' of the type invented by Wedgwood. Although it is difficult to know exactly the temperatures reached, Hall, fortunately, noted that 22 on his scale corresponded to the melting point of silver, which allows a very rough guess to be made of the temperatures actually reached. These must often have considerably exceeded 1000° C.

Not content with recording temperature, Hall devised a means of determining the pressure developed during his experiments, based, he states, on the means used by Count Rumford to measure the explosive force of gunpowder. The apparatus was crude by present-day standards (figure 8), but he could obtain approximate figures. The pressure was deduced from the weight lifted by the expansion of the gases generated in his compression tubes during an experiment. The maximum he recorded was 270 atmospheres; generally the figure was much less. This, he calculated, was equivalent to a column of sea water 8896 ft (2720 m) in height. He was, of course, interested in finding out the pressure exerted on rocks buried under a great depth of sea water or a thick cover of rock.

Hall accomplished most of his aims. He certainly established that calcium carbonate is not dissociated when heated under pressure, and from a geological point of view this was much the most important of his achievements. He also seems to have succeeded in turning powdered chalk or limestone into something very like a true crystalline marble, throwing light on the conditions of formation of certain naturally occurring marbles. Whether or not he really succeeded in melting calcium carbonate has been questioned. It is impossible to be sure from his descriptions, because, in those experiments in which he claims success, he does not mention having applied any test, such as effervescence with acid, which might have established the nature of the melt produced. Even with the resources of a modern laboratory it is not easy to melt calcium carbonate, and there is no unquestioned record of its having been done before 1912, when it was effected by H. E. Boeke, at a temperature of 1289° C and a pressure of 110 atmospheres [6]. It seems just possible that the necessary conditions may have been attained in some of Hall's experiments, but we shall never know for certain. What can be said, however, without fear of contradiction, is that what Hall did achieve was a very remarkable performance for an eighteenth-century amateur scientist, and of lasting value for geological science. He also deserves to rank as a distinguished pioneer in the history of physical chemistry.

Hall's results attracted much attention at the time, both in Great Britain and abroad, and they have often been quoted since. A French edition of his compression experiments was published in Geneva in 1807. It was translated by Marc-Auguste Pictet, a prominent Swiss scientist, who compared Hall with Lavoisier in a laudatory preface [7].

After the completion of the experiments described above, Hall continued to study geology and published several geological papers of interest. It is also worth remembering that his work shows him to have been not only an experimental geologist but a capable field geologist. He was elected a Fellow of the Royal Society of London in 1806, and in 1812 became the second President of the Royal Society of Edinburgh, an office he held until 1820. He died on 23rd June 1832 in Edinburgh, after some years of illness.

ACKNOWLEDGMENTS

Thanks are due to the Director of the Geological Survey and Museum, London, for permission to reproduce the illustrations of Hall's original apparatus, in the possession of the Museum; and to the Council of the Royal Society of Edinburgh for allowing reproduction of the portrait of Hall, which is in their possession, and the remaining illustrations taken from the accounts of his experiments in Volumes V and VI of the Society's 'Transactions'.

REFERENCES

[1] MACGREGOR, M. *Endeavour*, **6**, 109, 1947.
[2] MCKIE, D. *Notes roy. Soc. Lond.*, **7**, 13, 1949.
[3] HALL, Sir James. *Trans. roy. Soc. Edinb.*, **5**, 43, 1805.
[4] KEIR, J. *Phil. Trans.*, **66**, 530, 1776.
[5] HALL, Sir James. *Trans. roy. Soc. Edinb.*, **6**, 71, 1812.
[6] CLARKE, F. W. *Bull.* 770 *U.S. geol. Surv.* (fifth edition), 563, 1924.
[7] PICTET, M. A. 'Description d'une suite d'expériences qui montrent comment la compression peut modifier l'action de la chaleur', p. ix. Geneva. 1807.

Surface patterns in Boreal peatland

H. SJÖRS

The Boreal peatlands—that is, the bogs and fens in the zone around the North Pole where drained ground is mainly covered by coniferous forest—show very striking patterns. The origin of these patterns is not yet clear, and will become so only after extensive research. The present state of the study of possible origins is described and illustrated in this article.

It has been known for a long time that the northern bogs and fens show clearly defined patterns in which firm, dry areas alternate with soft, highly waterlogged ones. The origin of these striking patterns is still, however, unclear. There are a number of reasons for this. The areas are not always easily accessible on foot, as Linnaeus vividly recorded [1], and correlation of the work of different scientists has been hampered by differences of language. Aerial travel and aerial photography have, however, vigorously stimulated the studies—the areas can be more easily reached and the patterns more clearly seen—and a survey of the present state of research is now appropriate.

Analysis of local and regional distribution of the peatland patterns can serve as a starting point for a detailed study of their hydrotopography and hydrodynamics, but the problems are not purely geomorphological; an understanding of the state, development, and origin of the structures forming the patterns must be based on ecological research, starting from the present vegetation and its environmental conditions, and working backwards through detailed peat stratigraphy.

THE BOREAL ZONE

The Boreal zone is defined as the almost circumpolar zone in the north where drained grounds are covered by prevailingly coniferous forest [2]. Although patterned peatlands are not confined to this zone, they are far more common there than elsewhere. The Boreal forest, often referred to as Taiga, reaches from Trondheim to the Sea of Okhotsk, and from Alaska to Newfoundland (figure 13). Oceanic coniferous forests, such as the pinewoods of Scotland and westernmost Norway, and the Pacific forests, are not included, nor are other forests of non-Boreal conifers, of which there is a great variety. It is possible to subdivide the Boreal zone on a basis of the distribution of forest trees and other species, but a zonal subdivision is of greater general and ecological value. Apart from the transitional zone to the south, which in turn passes either into regions of deciduous forest or into steppes, the Boreal zone may conveniently be divided into main Boreal and sub-Arctic subzones. The former has been further subdivided, for instance by the Russians, into 'southern Taiga' and 'central Taiga' [3]. The sub-Arctic (the 'northern Taiga') is separated from the treeless Arctic by a usually narrow transition of woodland tundra. The zonation due to differences of altitude in mountainous parts of the Boreal zone roughly parallels zonation by latitude. However, the problem of equivalence is somewhat complicated by the presence of more or less subalpine birchwoods, particularly in Fennoscandia and Kamchatka.

The zones and subzones are closely related to the duration of the growth season, but they are little related to winter cold. All over the Boreal zone the winter temperatures are below freezing point, and seasonal soil frost is general. On the other hand, persistent strata of frozen soil, 'permafrost', are found only in the most continental parts of the Boreal zone, where the winter temperatures are exceedingly low. Some of the structures and patterns in peatland are evidently closely related to the type of soil frost, but others seem to be independent of freezing phenomena; the patterns in permafrost areas must be considered separately.

The distribution of vegetation shows no consistent relationship to that of permafrost [4]. Thus, in Europe and Labrador, discontinuous permafrost occurs in parts of the sub-Arctic. In other parts of Canada and in western Siberia some discontinuous permafrost is found as far south as the main Boreal subzone, but most of it is in the sub-Arctic; in the more continental parts, permafrost becomes continuous some distance south of the tree-line. In central and eastern Siberia, continuous and extremely deep permafrost is found throughout the Boreal zone, and even further south.

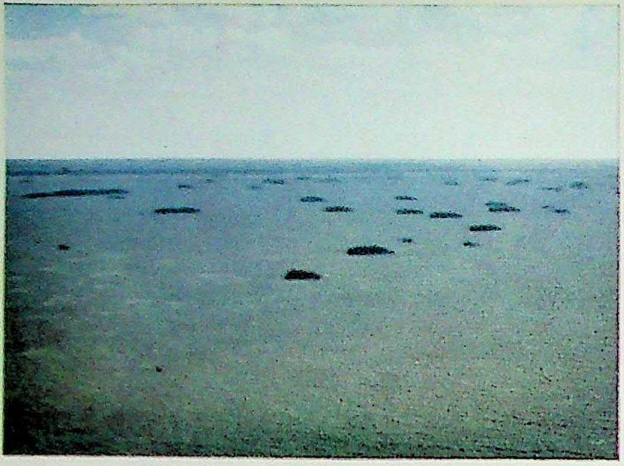


FIGURE 1 – *Expanse of minerotrophic peatland (fen). Small islands with a dense cover of black spruce* (Picea mariana). *All photographs on this page are from the Hudson Bay lowland in northern Ontario.*

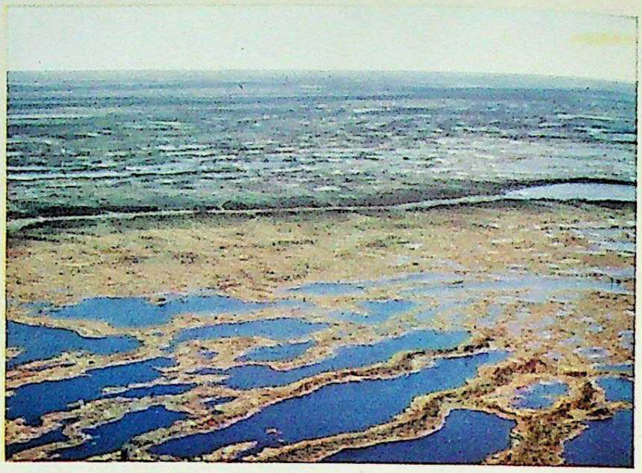


FIGURE 2 – *Ombrotrophic peatland (raised bogs) with numerous large bog-pools. A brook runs between two bogs (across the picture) and is expanded into a beaver pond (right).*

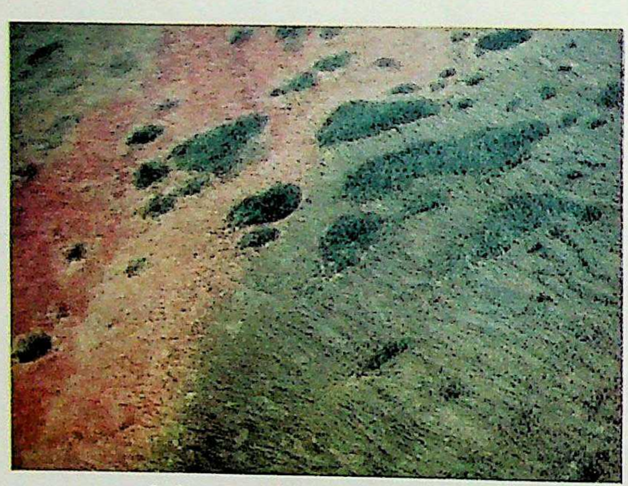


FIGURE 3 – *Left part ombrotrophic, colour mainly from* Sphagnum rubellum. *Right part minerotrophic, with pattern at right angles to slope, green colour partly from dwarf larch trees. Black spruce islands in centre.*

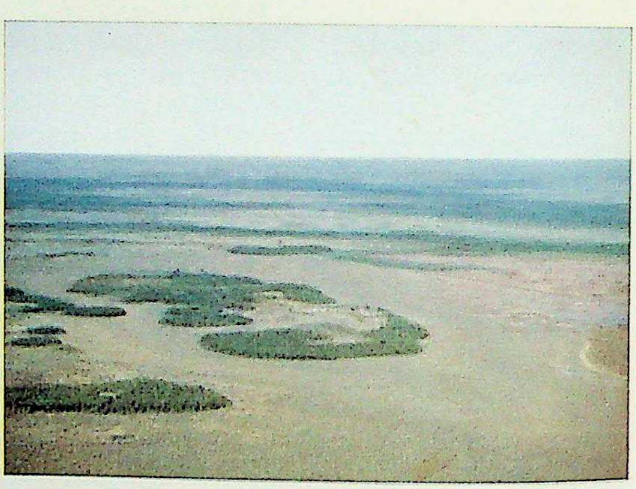


FIGURE 4 – *Large, circular, partly ring-shaped black spruce islands in a matrix of minerotrophic peatland largely covered by low larch trees* (Larix laricina).

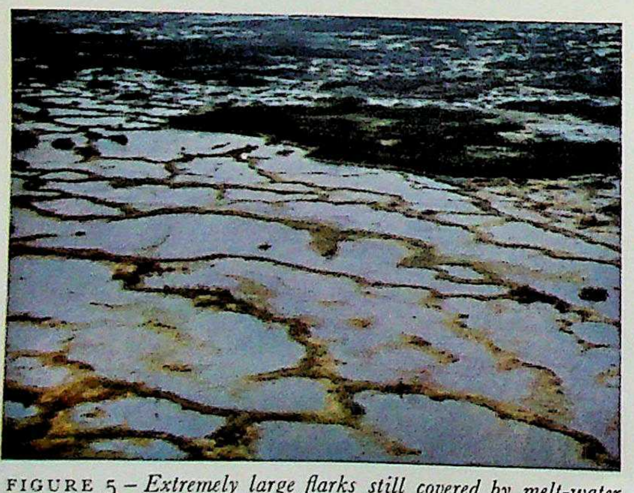


FIGURE 5 – *Extremely large flarks still covered by melt-water and separated by elevated ridges, partly with black spruce trees. This pattern is reminiscent of a rice-field.*

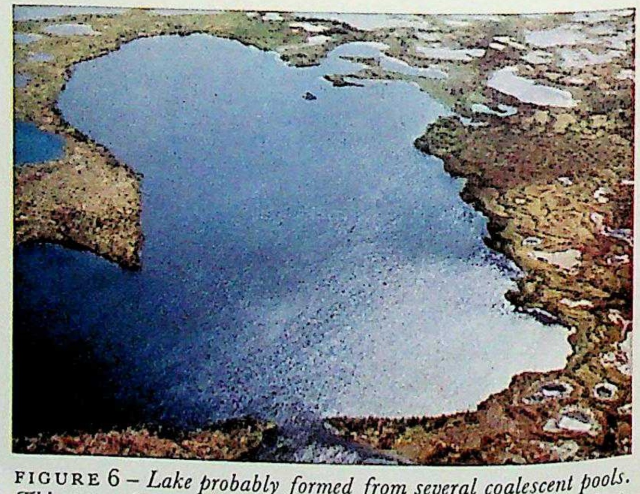


FIGURE 6 – *Lake probably formed from several coalescent pools. This type of pattern is found in almost flat areas which constitute water divides between rivers.*

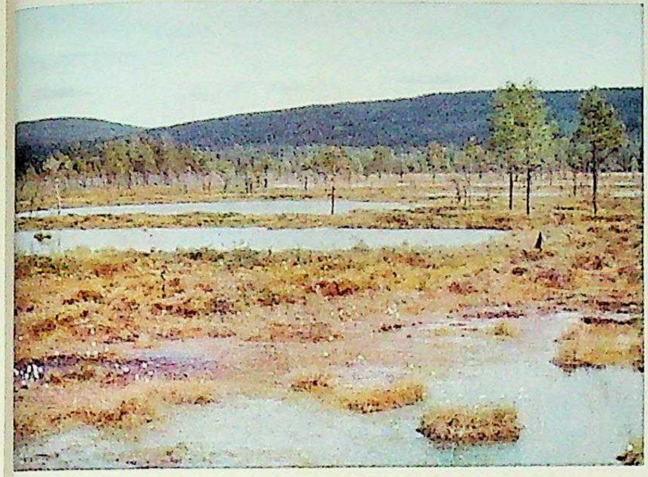


FIGURE 7 – *Ombrotrophic peatland with bog-pools. Scots pine; dominants* Calluna, Betula nana, *and* Eriophorum vaginatum; *at water's edge* Scheuchzeria palustris *and* Carex limosa. *Tisjökölen, Central Swedish uplands.*

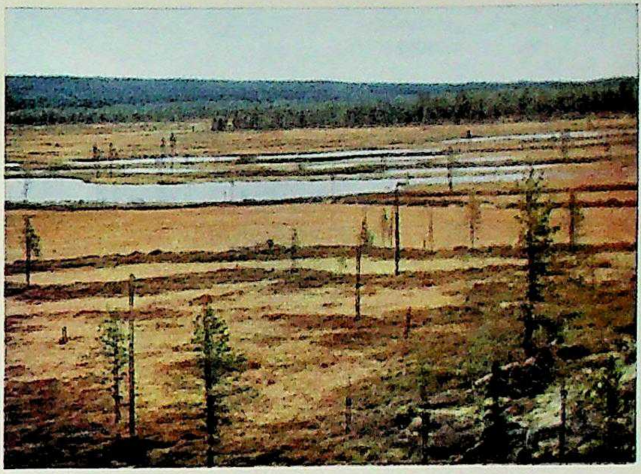


FIGURE 8 – *Ombrotrophic peatland with elevated ridges and a few bog-pools. Between ridges chiefly* Trichophorum caespitosum. *Foreground sandy, with* Cladonia *lichens and pine. Near Lofsdalen, Central Swedish uplands.*

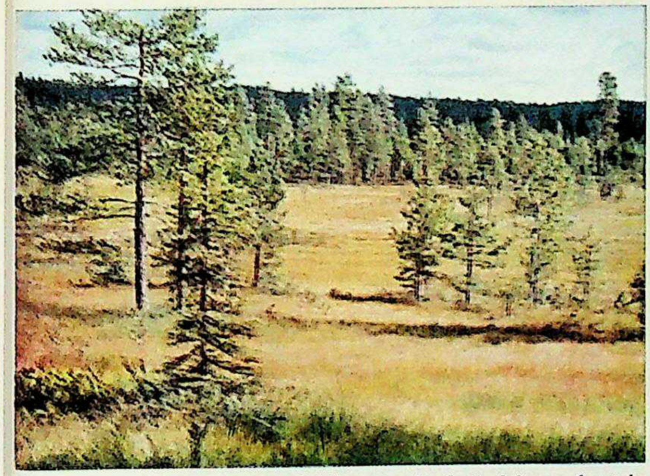


FIGURE 9 – *Sloping area, fairly strongly flushed by moderately minerotrophic water.* Molinia coerulea *in foreground. Sävsnäs, south part of Central Swedish uplands.*

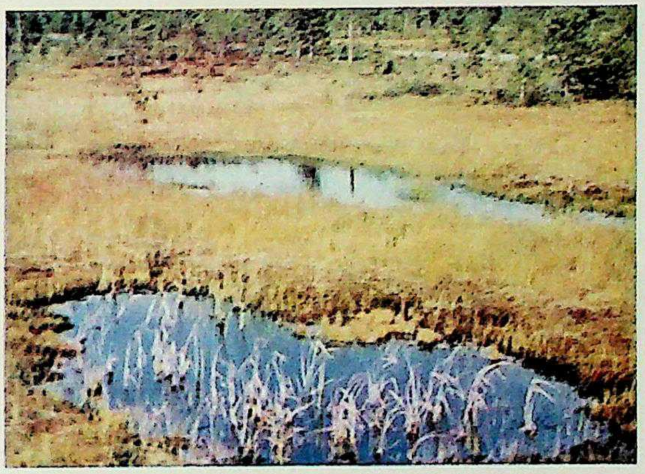


FIGURE 10 – *Two flarks separated by a strip of firm, but not much elevated, mat with* Carex *and* Molinia. *In the flarks* Carex rostrata, C. livida, *and* Rhynchospora alba. *Downhill of area shown in figure* 9.

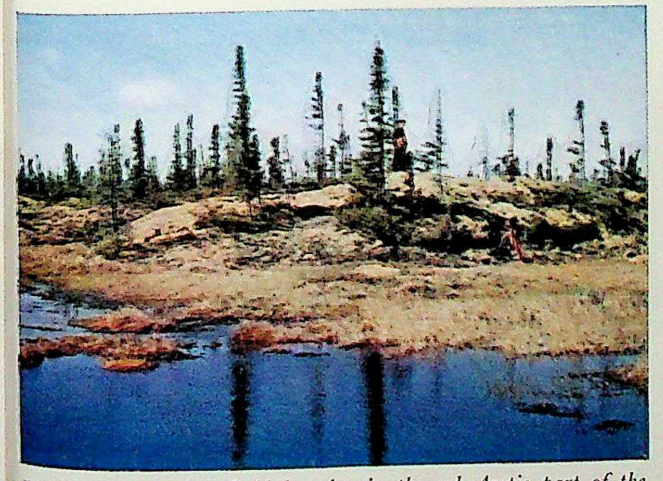


FIGURE 11 – *A 9 ft high palsa in the sub-Arctic part of the Hudson Bay lowland. Wind-worn trees of black spruce; carpets of lichens (largely* Cetraria nivalis) *indicative of lack of snow cover.*

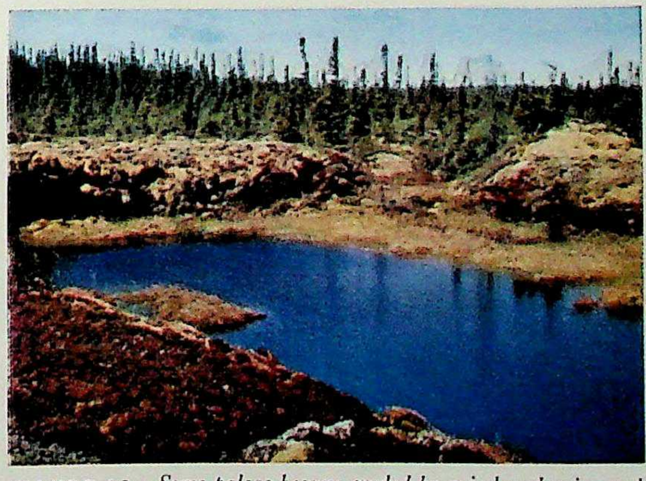


FIGURE 12 – *Some palsas become eroded by wind and rain, and then the ice core begins to melt. The pool is probably the site of a totally collapsed palsa. Same area as figure* 11.

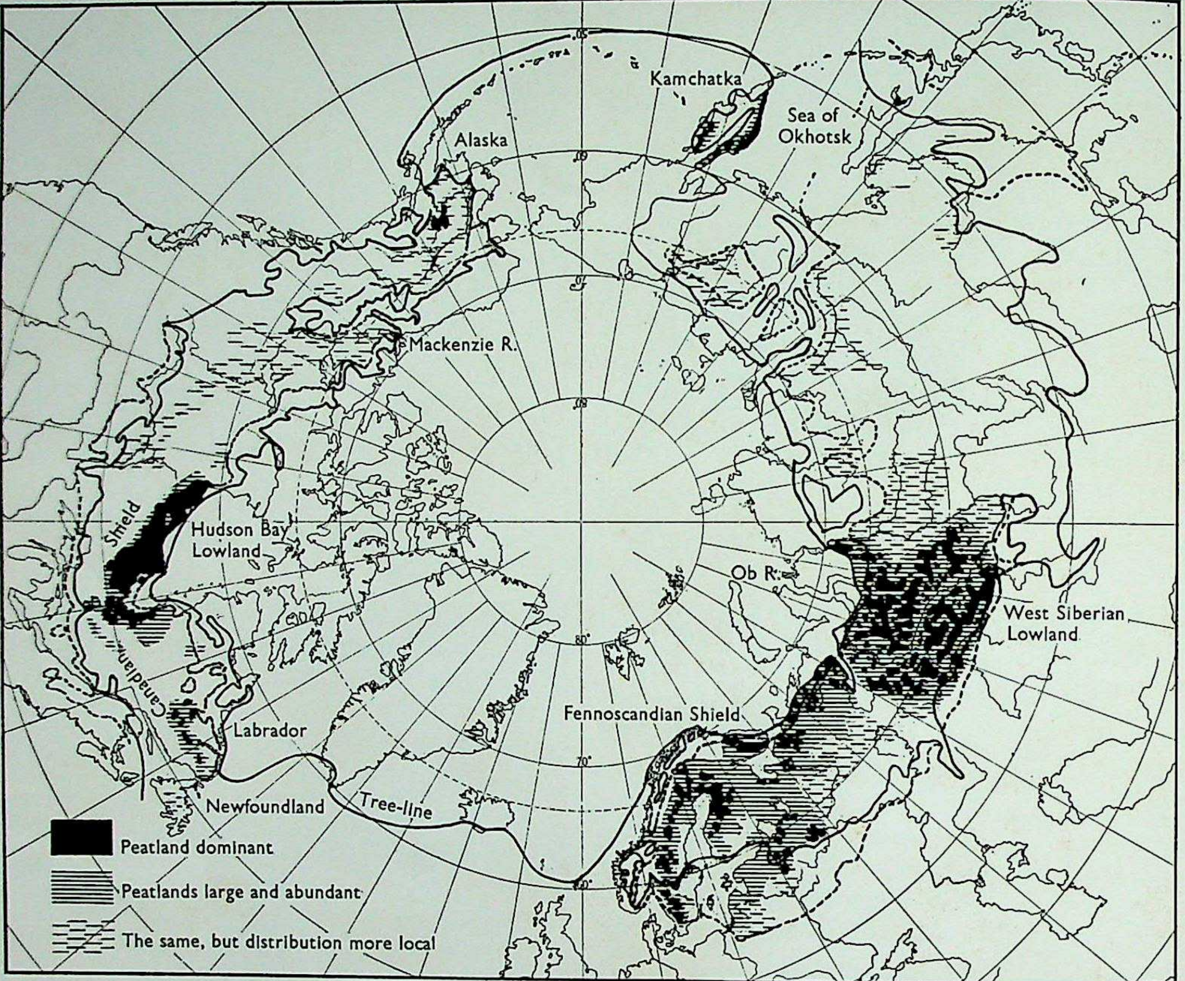

Kamchatka
Sea of Okhotsk
Alaska
Mackenzie R.
Shield
Hudson Bay Lowland
Canadian
Ob R.
West Siberian Lowland
Fennoscandian Shield
Labrador
Newfoundland
Tree-line
▇ Peatland dominant
▤ Peatlands large and abundant
▤ The same, but distribution more local

FIGURE 13 – *Area of abundant peatland in the Boreal zone. South of, or below, the tree-line, woodland tundra or subalpine areas extend to the northern broken line. South of the Boreal zone, transitional regions reach to the southern broken lines.*

The Boreal zone, like other biotic regions, is far from uniform in its vegetation. Boreal forests, some of which grow on peat, are less variable than more southern ones, but they still display a considerable diversity due not only to climate, but also to state of development, soil and water relations, and human influence. The most common types of unwooded areas are open bogs and fens, but their abundance is likely to be underestimated by an earthbound traveller.

PERCENTAGE OF PEATLAND

By far the greater part of the total peat area of the world is within the Boreal zone. However, because of variations in the relief of the land within this zone, the percentage covered by peat varies considerably. South of Hudson Bay there is a continuous lowland of about the same area as Great Britain. This lowland, with almost the only exception of lakes and river-beds, is completely covered by wet peat [5]. In other parts of Boreal America the peat deposits are less extensive, but there are considerable areas covered by peat, for instance in Alaska [6] and in Labrador [7].

The distribution of peat in Russia is uneven [8]. A large extent of western Siberia surrounding the Ob and Irtysh rivers is almost wholly covered by peatland, except along the rivers themselves and their tributaries. This is the largest peat area in the world. Some parts of Kamchatka and northern European Russia are also very rich in peat.

About one-third of Finland is covered by peat [9], the highest percentage for any country. In certain parts of Finland the percentage of land covered by peat reaches 60–70 per cent, and Sweden also has some areas where it covers more than half the land. About 20 per cent of the Boreal zone in Sweden lies under peat. Individual peat areas in Sweden and Finland are rarely very large, because of the lack of unbroken plains, but a few are quite extensive. Thus, the largest continuous peat area in Lappland, the Sjaunja, covers approximately 150 square miles.

The type of peatland, as well as the area, depends on the topography of the land, and also on the climate. It is perhaps true that the relative importance of climate has been overestimated—it is now clear that highly different types of peatland can exist, in the Boreal zone and elsewhere, in areas with practically identical climates.

THE NUTRITION OF PEAT VEGETATION

Two sources of mineral nutrients are available to peat, the atmosphere and the soil. The former obviously benefits all peatland, chiefly by means of the precipitation and dust that fall on them. The terrestrial supply, on the contrary, is very variable, and some peatlands, or parts of them, do not receive any at all. Areas dependent exclusively on nutrients from above are termed 'ombrotrophic' [10], that is, 'nourished by rain.' The sites that, in addition, receive terrestrial mineral nutrition are termed 'minerotrophic,' indicating that nutrients are brought to the peat by water that has previously extracted them from a mineral soil.

In oceanic areas, such as western Britain and western Ireland, ombrotrophic nutrition is only slightly inferior to weak minerotrophic nutrition, and little difference in vegetation results. The Boreal zone is largely subcontinental or continental, and ombrotrophic nutrition is much poorer over these areas, particularly with regard to the salts derived from the sea. With decreasing concentrations of mineral nutrients in the precipitation, the nutritional differences between ombrotrophic and minerotrophic peatlands naturally tend to become more marked.

On the other hand, Boreal soils and 'run-off' are such that much of the terrestrial water received by peatland has been in contact only with very poor and leached soils, or has been diluted by large quantities of almost completely ombrotrophic water. As a result, Boreal lands contain much peatland that is only weakly or even doubtfully minerotrophic. This is particularly so within the areas resting on the two large shields of hard Archaean or Pre-Cambrian rocks, the Canadian Shield and Fennoscandia.

Other areas show richer minerotrophic nutrition. Where fragments of limestone are represented in the soils, minerotrophic sites will be strongly enriched in calcium (but often not in other nutrients), and calcareous alkaline, or at least near-neutral, peats are formed. Elsewhere most, though not all, minerotrophic peats are more or less acid, some of them nearly as much so as ombrotrophic peats.

However, acidity is not a simple function of mineral supply. Among other factors influencing acidity, the rate of peat formation seems to be important. In waterlogged sites, a low rate usually results in less acid conditions.

The plant growth-rate and the organic production by photosynthesis are slower in the north, with its short growth-season; in addition the rate of decay of the peat in waterlogged sites seems to be faster in the north. The higher solubility of oxygen in water at low temperatures may be of importance, but most probably the vigorous growth of algae in these waterlogged sites is the main factor. The algae compete strongly with the peat-forming mosses, and they also produce much oxygen by photosynthesis, so that the organic material formed decays easily and no new peat accumulates. Moreover, remains of other plants, as well as existing peat, will be affected by a very active type of wet-surface oxidation, which occurs in addition to the normal decay. It has been shown [11] that there has been no peat formed for 2000–3000 years in some waterlogged areas, called 'hollows' and 'flarks', in the central Swedish uplands.

PRINCIPAL PLANT COMMUNITIES

Much has been written on the composition and ecology of the vegetation of Boreal peatland [9, 10, 12]. There are only a few plant species that are able to grow under strictly ombrotrophic conditions, and these plants also enter most minerotrophic sites; the majority of peatland plants, when growing in the Boreal zone, are limited to minerotrophic sites. Even in weakly minerotrophic (and strongly acid) peat areas, the flora is considerably richer than in ombrotrophic areas, and includes more species of sedges (*Carex*), cotton-grasses (*Eriophorum*), and peat-mosses (*Sphagnum*), and also the buck-bean (*Menyanthes trifoliata*), several willows (*Salix*), and other species. Near-neutral peatland and, still more markedly, alkaline calcareous peatland, have so many plants that these localities have been termed 'rich fen' [10].

The difference in vegetation between ombrotrophic and minerotrophic areas affects the general colour of the peatland, and they may sometimes be distinguished in aerial colour-photographs (figure 3). On black-and-white photographs their identification is not possible from the tone, but the patterns reveal so much of the hydrography that ombrotrophic and minerotrophic vegetation is usually readily distinguished. Other variations exist in Boreal peatland vegetation [6, 7, 9, 12]. They are effects of, for example, geographical and climatic differences, or of degrees of firmness and wetness.

Typical of Boreal peatland are extremely wet and muddy areas with much algal vegetation but scanty macro-vegetation. They are frequently very extensive but nearly always sharply circumscribed, and are regularly covered by shallow water after the melting of the snow and during other wet periods. They form the most distinctive part of the patterns, and seem to be stable and persistent. This has been shown by pollen stratigraphy [11]. Only rarely do they become filled in with any vigorously peat-forming vegetation. The growth of algae and the 'corrosive" superficial oxidation may explain this lack of recolonization and formation of new peat.

The patterns that occur in Boreal peatlands fall into two groups, those where permafrost does not occur, and those where it does. These groups can be considered separately.

PATTERNS IN AREAS WITHOUT PERMAFROST

In ombrotrophic bogs, ancient hollows that are 'left behind' tend to become deeper in relation to the surrounding bog surface when the peat there grows in thickness; finally they become deep secondary bog-pools [11, 13]. Bog-pools, usually from two or three to about ten feet deep, occur in large numbers in some Boreal bogs, such as in Sweden [11, 13] (figures 7 and 8) and in Quebec and northern Ontario [5] (figure 2). They are not confined to Boreal bogs. Recently, fine examples, evidently with a similar history, have been described from the Silver Flowe in south-west Scotland [14]. Other examples of patterned peatland outside the Boreal zone exist in northern Minnesota [15]. Within the Boreal, by far the most extensive and striking patterning is in the minerotrophic peatland, because of the comparatively large amounts of surface and subsurface run-off.

The principal feature of the Boreal minerotrophic patterns is the 'flark' (in Swedish) or 'rimpi' (in Finnish) (figures 5, 10). The French-Canadian term 'mare' has also been proposed [16]. A typical flark is a limited, usually elongated, wet, and muddy area accessible only with difficulty on foot, because the network of root-stocks often has too coarse a mesh to support the explorer's weight. Each flark is exactly horizontal, even when it is several hundred metres in length. On sloping sites the flarks are narrow, being only a few metres wide, but in almost horizontal peatland they may be a hundred metres wide or more.

The flark vegetation is remarkably uniform all over the Boreal zone, because the principal plants are circumboreal: there will be species of *Carex*, such as *C. limosa*, *C. chordorrhiza*, and *C. livida*, and also *Eriophorum angustifolium*, *E. gracile*, *Menyanthes trifoliata*, *Scheuchzeria palustris*, *Equisetum fluviatile*, *Juncus stygius*, and a few other species. There is still no reasonable explanation why these plants, which all propagate by branching rhizomes, grow so sparsely in the flarks. Nutrition is poor, but seemingly not worse than in other parts of the peatland, where the vegetation grows densely. All these plants are well provided with internal aerating tissues and do not suffer from the lack of oxygen at root level.

The long axis of the flarks is always at right angles to the slope, that is, in the direction of the contours. This is, in fact, true of all patterning that is independent of permafrost or strong winds. On a slope, even if slight, the consecutive flarks form steps, as in a staircase. Some of the large groups of flarks have a striking resemblance to a terraced rice-field (figure 5). In addition, the flark complexes are very clearly limited to the parts of the peatland through which most of the water has to pass on its way downhill.

Both these conditions show that the flarks are primarily dependent on the water flow across the peatland. Several other proposed explanations are ruled out by these basic facts of orientation and situation. It is not likely that downhill movements or soil-creep are the cause, for the flarks would then be found mainly on strong slopes and would be absent from nearly horizontal areas. In fact they are not. On the contrary, the most conspicuous flarks are found on slopes that are much too small to allow for any peat-flow at all (figures 5, 10), and there are no flarks on the steeper slopes (figure 9).

There is evidence [11] that peat growth is slightly better between the flarks than within them. This is necessary for the seemingly precarious stability of the flark system. Another prerequisite is that the damming rims should not be eroded. Indeed, natural erosion is exceptional, being observed only on slopes unusually steep for flark formation. When it occurs it destroys the pattern completely.

The flarks tend to be very large nearer the sub-Arctic parts of the Boreal zone. Here the dams separating them are frequently, though not always, marked ridges, usually two to three feet high, with lichens, dwarf-shrubs, and occasional low trees growing on them. The reason for the increase in height seems to be the pressure of the ice, exerted horizontally during the freezing period in early winter [17]. Similar phenomena occur in lake ice and are known to pile up stones and soil to form a low ridge round the lake. The expansion of the ice is due to temperature changes; when the ice shrinks in cold weather, numerous narrow cracks are formed which fill with water and then freeze over almost instantly. When the temperature rises again, the ice will expand and exert a pressure on the edges of the basin. The still unfrozen ridges between the flarks will become elevated because of the pressure exerted on them by the expanding ice on the flarks. The ridges become frozen later than the flarks, and they remain frozen much longer in spring.

Even if the pressure of the ice is the cause of the upheaval of the ridges it is probably not the cause of flark formation, nor of patterning as a general phenomenon, as the flarks are also often separated by much lower rims. Although the hydrodynamic pseudo-stability is not a full explanation of the origin of the flark patterns, their relation to hydrodynamic conditions must be taken into account in any hypothesis that is formulated.

PATTERNS IN AREAS WITH PERMAFROST

The first permafrost to appear in the peatlands is located in certain dry hummocks, which through the action of permafrost become increasingly elevated. Several types of elevated and frozen peat mounds are known and are called, *inter alia*, spruce islands and palsas.

Spruce islands are found in Alaska and the interior of Canada (figure 1). At the southern margin of the area of discontinuous permafrost, where the climatic conditions for tree growth are still good, the islands are often quite extensive, though moderately high, and covered by unusually dense stands of black spruce (*Picea mariana*) (figure 4). This species is well adapted to grow on peaty permafrost hillocks. The amount of soil it needs for rooting, and its need of nutrients for growth, are both low; it endures high acidity and reproduces profusely by 'layering', that is by rooting branches that become embedded in litter or moss and lichen vegetation.

Further north, the permafrost hillocks support only scattered low trees (black spruce and *Larix laricina* in Canada). They are blown free of snow in winter and are fully exposed to the cold air. Similar mounds, often entirely treeless, also occur in the northernmost parts of Russia and Fennoscandia [9, 18], where they are known as palsas (figure 11). They are usually of small size but may be up to 23 ft in height in the extreme north of Sweden and Finland, and still higher in Russia. Between the palsas, freezing is only seasonal, and the ground is usually very wet in summer.

The interior of palsas consists of alternating sheets of frozen peat and pure ice [18], and it is believed that water accumulates on the growing ice-crystals. This water, derived from below or from the surroundings, adds to the bulk of the palsa. A palsa frequently becomes eroded and ultimately may entirely collapse, and is then replaced by a water-filled pit (figure 12).

Where permafrost is continuous, the palsas are replaced by less conspicuous, but very extensive, elevated peat-plateaux, showing frost polygons and other types of pattern, which are very different from the patterns in the Boreal peatlands that are without permafrost. Occasional high mounds of ice and frozen peat occur also in areas of continuous permafrost, provided that liquid water is available. Near the Arctic these 'pingos' are very imposing; the famous pingos in the delta of the Mackenzie River [19] reach the dimensions of the Albert Hall in London.

HISTORICAL EVIDENCE

Only a small fraction of the Boreal peatlands have originated from lakes filled in by vegetation. However, such spots probably often served as starting points for the spread of peatlands over mineral soils. Other early centres were probably the result of the activity of beavers, which were at one time very numerous in these areas. Their dams cause waterlogging on the sides of streams and small rivers, which kills off forest trees and introduces peat-forming vegetation.

Most of the Boreal peats rest directly on mineral soil that once carried forest or at least swampy woodland or scrub. In fact, it seems unlikely that much peatland existed in the Boreal in the early part of the post-Glacial period, except perhaps in non-glaciated areas. Having regard to the recent origin of Boreal peatland and of its plant communities (particularly of the acidophilous types), their great uniformity in widely separated areas is astonishing. It is likely that most of the species are very old, but their combination into plant communities must be comparatively recent, so far as the present Boreal zone is concerned. The Boreal peatlands spread rapidly in the middle post-Glacial period. Later, this process became slower, and there is now little danger, in Fennoscandia, of deforestation by swamping [20].

Most of the patterns seem to have been lacking during early phases of development of the peatland and have probably originated during more recent millennia. Pollen dating [11] shows that the patterns are about 2000–3000 years old in the central Swedish uplands, and similar ages have been found for sub-Alpine palsas [18] in northernmost Sweden; more precise data may now become available by radiocarbon dating.

Succession in peatland includes much more than the normal successions starting in a wet habitat that are described in ecological textbooks. Obviously the reverse trend is more frequent, in which the development starts from mineral soil and the site gradually becomes wetter, with some parts becoming flarks or even secondary pools. Indeed, a further development seems to occur occasionally where adjacent pools expand and join to form lakes, as in some parts of the Hudson Bay lowland (figure 6). Structures in areas with permafrost also show both building and degradation.

The development of the northern peatlands is hardly in accordance with a uniform 'climax' theory. Rather than leading to uniformity, the paludification and further changes have increasingly diversified the habitats and their vegetation. In particular, the formation of patterns contradicts any climax concept, even a more modest 'polyclimax' view.

A biological material, in itself tending to disperse at random, is forced into a pattern of high mechanical precision by the ecological action of external forces and internal transformation. The variation is endless in detail, yet there is a surprising approximation to what looks like design but, of course, is not. Needless to say, biologists are puzzled, but also fascinated.

REFERENCES

[1] Linnaeus, C. 'Iter lapponicum'. Several Swedish editions, all posthumous. English edition by Sir James Smith, 1811 ('Lachesis Lapponica'). See also James, W. O., *Endeavour*, **16**, 107, 1957.

[2] Hare, F. K. *Geog. Rev., New York*, **40**, 615, 1950; *Geog. Studies, London*, **1**, 4, 1954; *Canada, Dept. Mines Tech. Surv., Geog. Branch*, mem. **6**, 1959.

[3] Lavrenko, E. M. and Sochava, V. B. (Ed.). Растительный покров СССР (Vegetational cover of the USSR). 2 vols., being the text of Геоботаническая Карта СССР (Geobotanical Map of the USSR), by the same editors (1954). Academy of Sciences of the USSR, Moscow and Leningrad. 1956.

[4] Brown, R. J. E. *Arctic*, **13**, 163, 1960.

[5] Coombs, D. B. *Geog. Bull., Ottawa*, **6**, 1, 1954.
Hustich, I. *Acta Geog., Helsingf.*, **16**, Pt. 1, 1957.
Sjörs, H. *Arctic*, **12**, 2, 1959; *Bull. Nat. Mus. Can.*, **171**, 1, 1961; *ibid.* (In preparation).

[6] Drury, W. H. *Contr. Gray Herb. Harv.*, **178**, 1956.

[7] Allington K. R. *McGill Sub-Arctic Res. Pap., Montreal*, 7, 1959 (Thesis). See also [2].

[8] Galkina, E. A. In [3], Vol. II, p. 553.

[9] Ruuhijärvi, R. *Ann. bot. Soc. 'Vanamo', Helsingf.*, **31**, Pt. 1, 1960.

[10] Du Rietz, G. E. *Svensk bot. Tidskr.*, **43**, 274, 1949; *Vegetatio, Haag*, 5–6, 571, 1954.

[11] Lundqvist, G. *Sveriges geol. Unders.*, **Ca21**, 1951.

[12] Cajander, A. K. *Acta for. fenn.*, **2**, 1913.
Malmström, C. *Medd. Skogsförsöksanst. Stockh.*, **20**, 1923.
Kotilainen, M. J. *Wiss. Veröff. Finn. Moorkulturver.*, **7**, 1928; *ibid.*, **19**, 1951.
Paasio, I. *Acta. for. fenn.*, **39**, 1933.
Sjörs, H. *Ark. Bot.*, 33A, Pt. 6, 1946; *Acta phytogeogr. Suec.*, **21**, 1948; *Bot. Notiser*, 1950, 173, 1950; *Oikos*, **2**, 241, 1950.
Gorham, E. *Quart. Rev. Biol.*, **32**, 145, 1957.

[13] Granlund, E. *Sveriges geol. Unders.*, **C 373**, 1932.

[14] Ratcliffe, D. A. and Walker, D. *J. Ecol.*, **46**, 407, 1958.

[15] H. E. Wright and M. L. Heinselman. Unpublished information.

[16] Hamelin, L.-E. *Cahiers Géog., Québec*, **2**, 87, 1957.

[17] Auer, V. *Acta for. fenn.*, **12**, Pt. 2, 1920.

[18] Lundqvist, G. *Geol. Fören. Stockh. Förh.*, **73**, 209, 1951.

[19] Porsild, A. E. *Geog. Rev.*, **28**, 46, 1938.

[20] Malmström, C. *Medd. Skogsförsöksanst. Stockh.*, **26**, 1, 1931.

Positronium

F. F. HEYMANN

Positronium, a very short-lived complex of an electron and a positron, has some properties resembling those of hydrogen, and can in some respects be looked upon as the lightest of the elements. For example, it reacts with halogens to form halides akin to those of hydrogen, and for each line in its optical spectrum there is a corresponding one at half the wavelength in the spectrum of hydrogen. Positronium closely links physics and chemistry, requiring the techniques of both for its study. This article describes the history and present state of the research.

POSITRONS AND ELECTRONS

The Schrödinger wave equation, which specifies the wave-like behaviour of an elementary particle, is applicable only to cases in which the velocity of the particle is negligible in comparison with that of light. One of the most striking predictions that resulted when P.A.M. Dirac, in 1930, developed a wave equation in which proper account was taken of the consequences of the Special Theory of Relativity, was that of the existence of a positively charged electron. Two years later, a cloud chamber track of this particle, now known as the positron, was indeed observed by C. D. Anderson.

From Dirac's theory, and the theory of Quantum Electrodynamics, it was deduced, and has since been amply verified observationally, that electrons and positrons can be created in pairs in a variety of suitable circumstances, and, conversely, that an encounter between a positron and an electron may lead to the annihilation of both particles, the energy associated with their masses and motion being converted into electromagnetic radiation in the form of γ-rays.

The electron and the positron are both spinning particles. Each may be thought of as an electric charge rotating about its own axis in a top-like fashion. A natural consequence of this rotation is that both particles possess a magnetic field, that of a positron being directed oppositely to that of an electron spinning in the same sense. The inherent angular momentum of these particles is referred to as their spin.

Spin, being an angular momentum, is subject to certain well-established quantum rules. One important consequence of these rules is that, when an electron and a positron are close to one another, their spin axes can be only either parallel or antiparallel to one another. These two possible states of affairs are known as the triplet and singlet states respectively, and will be referred to frequently in this article.

By invoking the principle of conservation of angular momentum, it can be shown that when a positron and an electron annihilate one another in a head-on collision in the singlet state, all the initial energy is converted into two γ-rays emitted in directions 180° apart when measured in the C.M. co-ordinates (co-ordinates in which the centre of mass of the initial particles is at rest). If the particles move relatively slowly before the collision, which is so in all the cases to be discussed here, the energy involved is substantially just that associated with the rest-masses of the two particles. In numerical terms, each γ-ray will then have an energy of 0·511 MeV.

In the case of a head-on annihilating collision in the triplet state, it can be shown that three γ-rays must be produced. These three γ-rays are coplanar in the C.M. co-ordinates and the sum of their energies is 1·022 MeV, but the energy of any one of them is statistically distributed in a continuous spectrum as illustrated in figure 1. On the same diagram the singlet-annihilation radiation would be a line spectrum at 0·511 MeV.

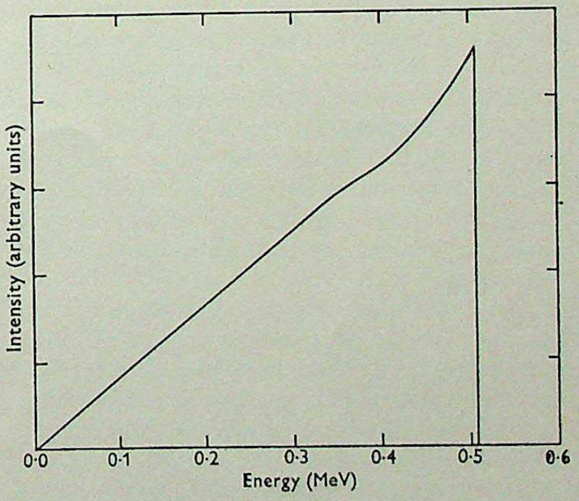


FIGURE 1 – *Triplet-annihilation spectrum* [5].

Glancing collisions hardly ever result in annihilation; the two cases described above are the only ones of practical importance.

The probability of a triplet-annihilating collision is only about 1/1120 of that of a singlet annihilation. This means that the number of triplet annihilations makes an almost negligible contribution to the rate at which positrons are annihilated in free collisions when moving through a cloud of electrons with randomly oriented spin-axes. Nevertheless, the smallness of the triplet-annihilation probability is the very feature which makes a study of positronium at all possible, as will be shown later.

THE NATURE OF POSITRONIUM

In the preceding discussion of the annihilation of a positron and an electron it has been tacitly assumed that the two particles meet in a free head-on collision leading to their disappearance. Although the ultimate fate of every positron is an act of annihilation with an electron, it is indeed fortunate for the enquiring mind that the path to death of a positron frequently occurs through a more devious route.

It is, for instance, possible for a positron moving in a gas to prise loose one of the electrons of a gas molecule and to remain attached to it. In classical language, this electron-positron pair rotates around their common centre of mass, held in dynamic equilibrium by a balance between their electrostatic attraction and the centrifugal force due to the rotary motion. It is tempting to describe such a structure as an atom; and there is a close analogy between it and the hydrogen atom in which an electron and a proton, nearly two thousand times heavier than the positron, rotate about their common centre of mass. The electron-positron pseudo-atom is what is known as positronium.

It should be stressed that this formation of bound states is by no means a rare phenomenon. It plays a major part in the study of positron annihilation. Experimental evidence is available to show that, typically, 30 per cent of all slow positrons in a gas form positronium atoms before annihilating.

Depending on the original relative orientations of the electron and positron spin-axes, positronium can be formed in either the triplet or the singlet state. From statistical arguments it can be shown that formation in the triplet state is three times as likely as in the singlet state.

Although positronium is a stable structure in the sense that it requires energy to separate the particles from one another, the positron and the electron nevertheless eventually suffer mutual annihilation, and have a predictable mean life. In the case of the singlet state, this is in fact about $1{\cdot}25 \times 10^{-10}$ seconds, with the usual inverse exponential distribution of individual lifetimes. As was mentioned previously, annihilation in the triplet state is less likely by a factor 1/1120, leading in this case to a mean life of about $1{\cdot}4 \times 10^{-7}$ seconds. These times may sound negligibly short, but they are quite long on an atomic scale. To be sure, the mean life of singlet positronium is too short to allow observations to be made on more than a few of its properties with present-day techniques; but 10^{-7} seconds is long enough to study many aspects of the physical and chemical behaviour of triplet positronium.

Positronium has a system of excited states, the energies of which can be calculated in exactly the same way as that of the hydrogen atom. In fact, apart from differences in the fine structure of the levels, primarily due to the differences in mass and magnetic moment of protons and positrons, the energy levels of positronium are similar to those of atomic hydrogen, with all energies scaled down by a factor which is almost exactly 2. The ionization energy—the energy required to just separate the positron and the electron from unexcited positronium—is consequently 6·8 eV, or half that of the hydrogen atom. Moreover, for each line in the optical spectrum of hydrogen there is a corresponding line in the positronium spectrum at double the wavelength.

QUENCHING OF TRIPLET POSITRONIUM

Once formed, the positronium atom will make random collisions with the gas molecules in which it moves, frequently leading to chemical and quasi-chemical processes. Evidence has, for instance, been cited for the existence of such chemical compounds as halides of positronium.

The effect usually observed as arising from interactions between positronium and gas molecules is an apparent reduction in the mean life of triplet positronium by processes collectively known as quenching collisions.

Probably the most frequently occurring quenching mechanism is one in which, during a collision between a positronium atom and a gas molecule, the positron annihilates directly with one of the electrons of the molecule in a singlet orientation. At first sight, one is led to suggest that this would yield a rate of annihilation comparable with free positrons moving in the same gas. In the case of positronium, however, the positron is screened to a large extent by the electron, resulting in a reduction of the annihilation rate by about one order of magnitude. In gases in which this is the major process, one would therefore expect to find a moderately small degree of quenching.

Two groups of gases are known to be markedly strong quenching agents. The first group, notably NO and NO_2, contains an odd number of electrons in the molecule. The process responsible for quenching in this case is thought to be one in which, during a collision, the electron of the triplet positronium is exchanged for the odd electron in the molecule, which may be in a suitable orientation to result in singlet positronium. The singlet state then annihilates rapidly with its own characteristic mean life of $1{\cdot}25 \times 10^{-10}$ seconds.

The second group of strongly quenching gases, namely the halogens, are thought to form chemical compounds such as positronium chloride, analogous to hydrochloric acid. In such a compound, instead of wandering freely from molecule to molecule, the positronium is bound to remain close to the chlorine electrons, thus greatly enhancing the probability of an annihilating collision.

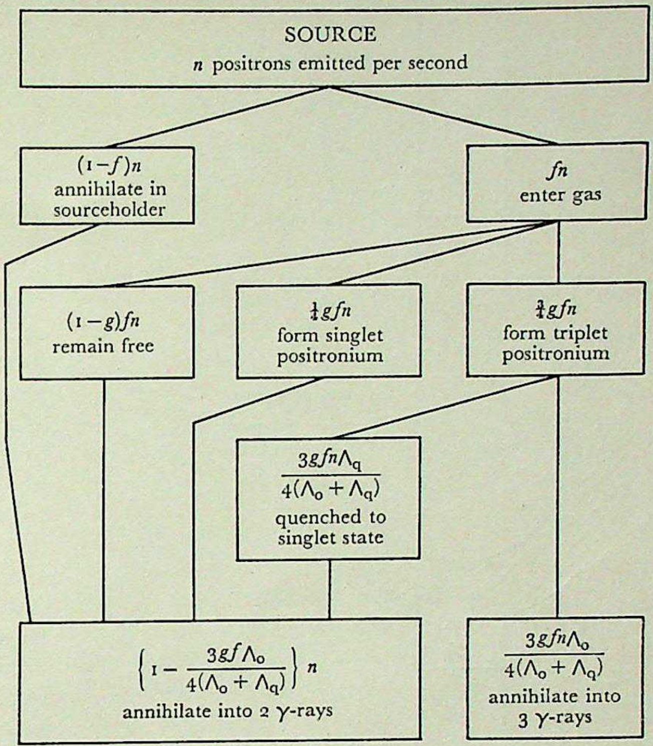

SOURCE
n positrons emitted per second
$(1-f)n$ annihilate in sourceholder
fn enter gas
$(1-g)fn$ remain free
$\frac{1}{4}gfn$ form singlet positronium
$\frac{3}{4}gfn$ form triplet positronium
$\frac{3gfn\Lambda_q}{4(\Lambda_o+\Lambda_q)}$ quenched to singlet state
$\left\{1-\frac{3gf\Lambda_o}{4(\Lambda_o+\Lambda_q)}\right\}n$ annihilate into 2 γ-rays
$\frac{3gfn\Lambda_o}{4(\Lambda_o+\Lambda_q)}$ annihilate into 3 γ-rays

FIGURE 2 – *Possible fates of positrons in a gas* [6].

FATE OF POSITRONS IN A GAS

The positrons utilized for studies on positronium are usually obtained by placing a radioactive source such as ^{22}Na, which decays by emitting an energetic positron, in the gas under observation. The positrons are rapidly slowed down by collisions with gas molecules until they move slowly enough either to annihilate or to form positronium. This slowing-down time is comparatively short (about 10^{-9} seconds) in comparison with both the mean life of triplet positronium and the mean life of positrons against annihilation in free collisions with electrons.

A chart summarizing the possible alternative courses of the history of a positron in a gas is shown in figure 2. In this diagram the factor f represents the fraction of the emitted positrons that manages to leave the source and enter the gas without annihilating with electrons of the material of the source or source-holder. The factor g represents the fraction of the positrons entering the gas that forms positronium.

The quantity Λ_o is the natural annihilation rate of triplet positronium, that is, the reciprocal of the mean life in the absence of quenching; while Λ_q is the quenching rate, that is, the reciprocal of the mean life if quenching were the only factor causing disappearance of triplet positronium.

DIRECT OBSERVATION OF POSITRONIUM LIFE IN A GAS

It is one of those rare fortunes of nature that the radioactive material commonly used for studying positronium emits a 1·36 MeV γ-ray practically simultaneously with the emission of the positron. The detection of this γ-ray thus provides a convenient time marker for registering the instant of birth of a positron. The instant of annihilation can be observed accurately by detecting any one of the annihilation γ-rays. A sketch of typical equipment used for observations of this kind is shown in figure 3. R is the radioactive source, placed inside a high-pressure vessel C. A scintillating crystal X_1 detects the passage of a 1·36 MeV γ-ray, thus defining the time of birth of a positron. This crystal is optically coupled to a photomultiplier P_1 which converts the light flash into an electrical impulse and amplifies it. A lead plug L prevents radiation from the source from reaching the crystal X_2, which can, however, respond to radiation produced by a positron annihilating in the gas. This crystal, in conjunction with the

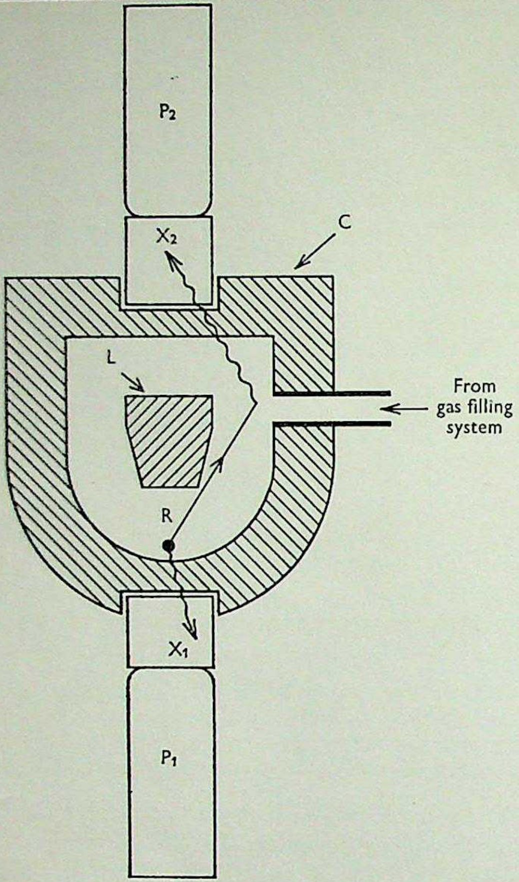


FIGURE 3 – *Apparatus for observing positron life.*

photomultiplier P_2, produces a pulse to define the instant of annihilation. The time intervals between individual 'birth' and 'death' pulses are measured and recorded electronically. With suitably fast electronic circuits it is possible to measure these time intervals to an accuracy of about 10^{-9} seconds, which is around 1 per cent of the mean life of triplet positronium.

In experiments of this kind, positrons that form singlet positronium immediately after slowing down annihilate so rapidly that, because of the finite time-resolution of the equipment, their lifetimes show up as effectively prompt coincidences between the birth and death counters. Positrons annihilating by free collisions with electrons and triplet positronium do, however, usually yield measurably long lives. This is illustrated by the typical delayed coincidence curve of figure 4. This curve can, for instance, be analysed into three components as shown, of which the slow exponential component, manifesting itself as a straight line on a logarithmic plot, contains all the information necessary to determine the mean life and hence its reciprocal, the annihilation rate.

Lifetimes arising from free annihilations may be distinguished experimentally from those due to triplet positronium by virtue of the fact that their annihilation rates differ in their dependence on pressure. In the case of free annihilation, positrons will collide with electrons of the gas molecules at a rate proportional to the mean electron-density. In other words, the annihilation rate under these circumstances will be proportional to the gas pressure. Results obtained from observations in oxygen (figure 5) illustrate this state of affairs. Oxygen is such a strong quenching agent that, even at small pressures, triplet positronium has such a short life that the only significant contribution to the long-lived component is that due to annihilation of free positrons. The curve of annihilation rate against pressure is accordingly a straight line passing through the origin, corresponding to a strict proportionality.

In contrast to this, gases of the Freon series, typified by CCl_2F_2, have the remarkable property of mopping up free positrons and causing them to annihilate rapidly, leaving triplet positronium as the only contributor to the long-lived component. In the absence of quenching, the mean life of triplet positronium is a quantity characteristic of the positronium atom alone, independent of gas pressure. This is indeed found to be the case in Freon, as shown in figure 5.

In practice, the long-lived component usually has to be ascribed to a mixture of free positron, and triplet-positronium, annihilations. The life curve under these circumstances will consist of a mixture of two exponentials. The two parts may be

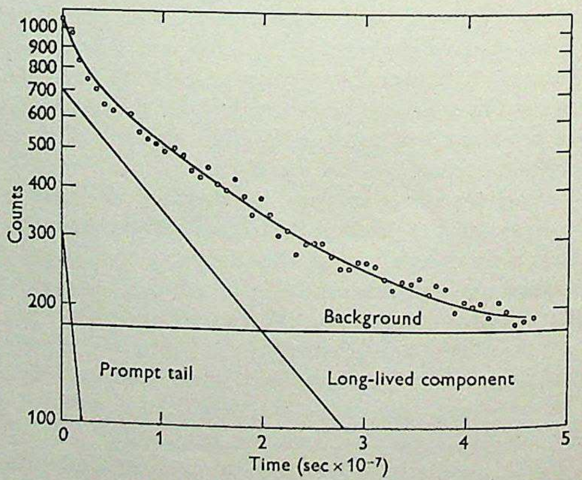


FIGURE 4 – *Delayed coincidence curve in Freon* [4].

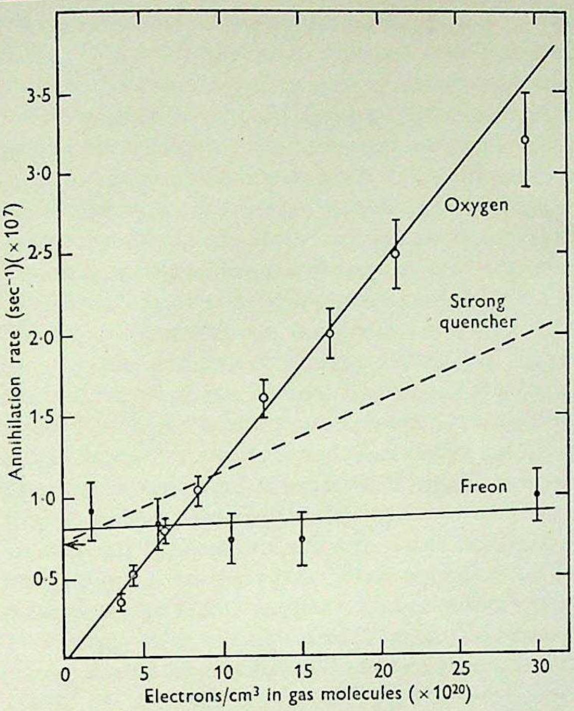


FIGURE 5 – *Pressure-dependence of annihilation rates in oxygen and Freon* [3]. *Arrow indicates theoretical triplet-rate* [5].

observed separately by adding either a small amount of Freon to the gas under observation in order to remove the free positrons, or a small amount of a strong quenching agent such as NO in order to remove the triplet positronium.

The third line in figure 5 illustrates the effect of a gas with a strong quenching action. The variation of lifetime with pressure is such that the annihilation rate consists of two additive components; a constant rate due to the natural mean life, with a quenching rate proportional to electron density superimposed on it.

γ-RAY-SPECTRUM OBSERVATIONS

As was shown earlier (figure 1), the annihilation γ-rays for triplet annihilations have a continuous energy-spectrum, while the spectrum of singlet annihilation consists of a sharp line. This difference between the two spectra is used to determine the effective concentration of triplet positronium in a gas sample.

When a beam of mono-energetic γ-rays, such as those resulting from singlet annihilations, impinges on a scintillating crystal such as sodium iodide, the resulting electrical pulses from a photomultiplier coupled to the crystal have a height distribution as shown in figure 6, curve (i). The relatively sharp peak is produced by γ-rays suffering photoelectric absorption in the crystal, in which case the magnitude of the light pulse is directly proportional to the energy of the γ-ray. The long plateau to the left arises largely from those γ-rays that are scattered by Compton collisions with electrons, in which case a smaller and variable fraction of the γ-ray energy is converted into light. Because of this complex response of a scintillation spectrometer to even a mono-energetic beam of γ-rays, it is in general not practicable to retrieve detailed information about the energy distribution in a continuous range of γ-ray energies from the resulting pulse-height distribution. Nevertheless, if such a unit is allowed to view the γ-rays produced in a chamber where both singlet and triplet annihilations are occurring, the resulting pulse-height spectrum is modified to that shown in figure 6, curve (ii). The increase in the valley height, shown by the arrows, is a sensitive indication of the number of three-quantum annihilations.

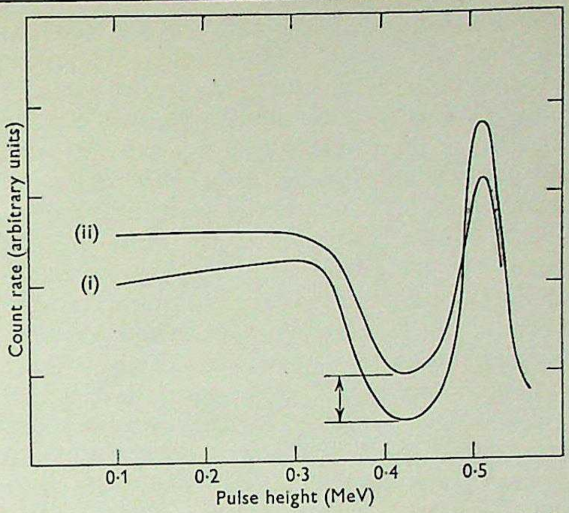


FIGURE 6 – *Scintillation pulse-height spectra from annihilation radiation.*

What is perhaps the simplest demonstration of the existence of triplet positronium may therefore be made by comparing the valley height arising from annihilation radiation in a chamber containing a positron source and a weakly quenching gas with that which results when a small admixture of NO is introduced into the chamber. In the latter case the triplet state is strongly quenched, so that practically all the annihilation γ-rays are in the form of two-quantum radiation.

OTHER DETECTION METHODS

The remaining important methods that have been used for identifying the origin of annihilation radiation make use of the facts that the multiplicity and the angular distributions of triplet radiation differ from those of singlet radiation.

For instance, two scintillation-counters viewing a chamber from opposite directions will give coincident output pulses as a result of singlet annihilations. This has been used indirectly to observe triplet positronium, by observing the decrease in the intensity of two-quantum radiation when conditions in the chamber are altered to result in more three-quantum annihilation, such as by doing successive observations with and without small admixtures of NO to the gas in the chamber.

The emission of three simultaneous γ-rays in triplet annihilations has been verified directly by placing three scintillation-counters with their centres coplanar with the centre of the positronium chamber, in positions such that no straight line through any two crystals intersects the chamber. Threefold coincidence-pulses between these counters then indicate the occurrence of a triplet annihilation. By this means the coplanarity of the radiation has been verified by observing the reduction in counting rate when one of the counters is displaced from the coplanar position.

Some of the more sophisticated observations on positronium have used more than one of the detection techniques simultaneously. For instance, spectral measurements on pure triplet positronium have been made in which the pulse-height distribution of one of the scintillation counters in a triple counter arrangement was analysed, counting only those pulses accompanied by coincident pulses from the other two.

MAGNETIC QUENCHING

When a triplet positronium atom is placed in a magnetic field, the interaction between the field and the magnetic moments of the electron and the positron may cause the spin axis of one or the other particle to be reversed in direction for part of the time. In the absence of other effects there is then a distinct probability, depending on the strength of the magnetic field, of finding singlet positronium in what was initially a sample of triplet positronium. In the language of quantum mechanics, the presence of a magnetic field causes mixing to occur between the amplitudes of the state with unit angular momentum and a magnetic quantum number of zero (a certain fraction of the triplet positronium) and of that with zero angular momentum (singlet positronium). During that fraction of the time in which the original triplet atom behaves like a singlet atom there is, of course, a greatly enhanced probability that it will annihilate rapidly, thus effectively leading to a reduced mean life for the triplet state. An equivalent way of stating this is to say that, in a magnetic field, triplet positronium can annihilate in the two-quantum mode characteristic of the singlet state, the magnetic field itself absorbing any unbalance in angular momentum necessary to satisfy the conservation laws.

If a positron source is placed in a gas sample, the concentration of triplet positronium in it will reach a dynamic equilibrium when the rate of formation from newly emitted positrons is balanced by the rate of loss due to annihilation. A reduction in the mean life will then lead to a reduction in the concentration. This fact has been used to observe magnetic quenching, using the valley-height method of measuring the intensity of the triplet-annihilation spectrum. A typical curve plotting the rate of triplet annihilations against magnetic field is shown in figure 7. The observations are entirely in agreement with detailed theoretical calculations.

ENERGY LEVELS AND OPTICAL SPECTRUM

If triplet positronium were formed in an excited state, there would be a distinct probability of emission of one of its characteristic optical spectral lines if the mean life against de-excitation by optical emission were short in comparison with the mean life against direct annihilation in the excited state. In particular, this is the case for that line analogous to the Lyman α line of atomic hydrogen. In the positronium spectrum, this line has a wavelength of about 2430 Å. Although this is twice the wavelength of the hydrogen line, it is

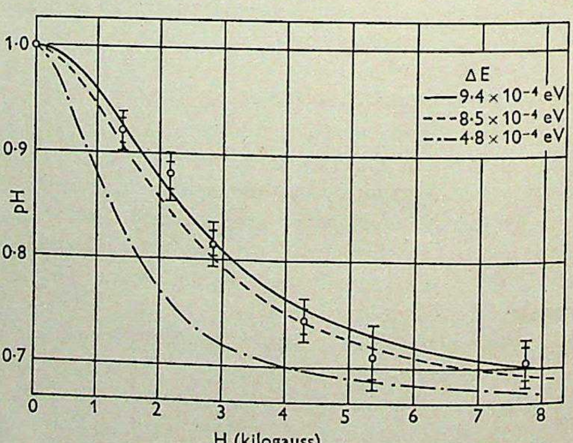


FIGURE 7 – *Quenching of triplet state by a magnetic field* [2]. *The various curves correspond to different theoretical interpretations.*

still well in the ultra-violet region, which makes for some complication in its detection. In addition, the concentrations of positronium obtainable in practice are extremely small in comparison with gas-concentrations typically used for spectroscopic work, so that one would not expect to detect more than one photon every few seconds in a typical laboratory experiment. Although a number of attempts have been made to locate this line, no completely unambiguous result has yet been obtained.

In spite of the fact that no optical line in the positronium spectrum has yet been observed, one is confronted with the experimental situation that certain details of the fine structure of the lines due to the Zeeman effect have been observed.

In a magnetic field, spectral lines in general are found to be split into components differing very slightly in wavelength. This is because the magnetic field causes various energy-levels of the atom to be split into close-lying components, corresponding to different values of the magnetic quantum number. By irradiating the atoms with a radio-frequency electromagnetic field superimposed on the static magnetic field, it is possible to induce transitions between these sub-levels when the frequency is that corresponding to the differences in energy involved.

In a noteworthy set of experiments of M. Deutsch and S. C. Brown, this process was in fact observed in the following manner. A positronium chamber in the shape of a microwave-resonant cavity was placed in a static magnetic field strong enough to ensure practically complete magnetic quenching. In these circumstances, as was mentioned earlier, almost all the triplet positronium with magnetic quantum number zero would be quenched. A radio-frequency field was then applied to the cavity, whereupon further quenching was observed, at that value of the static field which caused the correct level-splitting to enable the radio-frequency field to induce transitions from the sub-levels with magnetic quantum number ± 1 to the sub-level with magnetic quantum number zero. This additional quenching, plotted against magnetic field, is shown in figure 8.

The position of the minimum in this curve agrees to a high accuracy with the value calculated from quantum electrodynamics when a number of refined theoretical arguments are included.

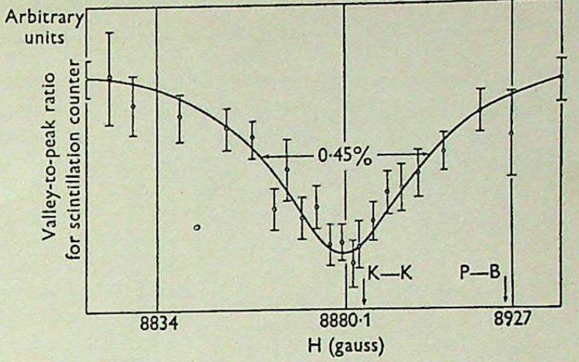


FIGURE 8 – *Quenching of* $m = \pm 1$ *level by radio-frequency field* [1]. *P—B is expected minimum position according to first-order electrodynamic theory. K—K is expected minimum position including higher-order corrections.*

This was a remarkable experiment by any standards, and is to be compared with the famous Lamb-Retherford experiment on the ground-state-level splitting of the hydrogen atom, as being one of the few precision experiments that have been performed to prove the correctness of certain abstruse higher-order effects in electrodynamics.

CONCLUSION

In an article of this length it has not been possible to do more than touch on some of the more salient features of the study of positronium. The author does, however, hope to have indicated that the detailed study of the fate of a positron in matter has proved to be a subject with many more facets of interest than was originally expected from the simple Dirac theory of annihilation by head-on collisions between positrons and electrons, in so far as this quasi-atomic complex of a transitory nature known as positronium may be formed.

The fascination of the experimental study of the behaviour of positronium lies perhaps in the particular way that it allows the nuclear physicist to contribute to chemistry and spectroscopy, using his now traditional techniques of scintillation counting. In this field microchemistry has in one sense been carried to its ultimate limit, in that the reaction times of individually reacting atoms have been observed.

REFERENCES

[1] DEUTSCH, M. and BROWN, S. C. *Phys. Rev.*, 85, 1047, 1952.
[2] DEUTSCH, M. and DULIT, E. *Ibid.*, 84, 601, 1951.
[3] DEUTSCH, M., DULIT, E., and PRIMAKOFF. *M.I.T.L.N.S.E. Prog. Rep.*, 131, 31st May, 1951.
[4] DUFF, B. G. Private communication.
[5] ORE, A. and POWELL, J. L. *Phys. Rev.*, 75, 1696, 1949.
[6] OSMON, P. E. Thesis, University of London, 1960.

Book reviews

THEORETICAL PHYSICS

Theoretical Physics in the Twentieth Century: a memorial volume to Wolfgang Pauli, *edited by M. Fierz and V. F. Weisskopf. Pp. x+328. Interscience Publishers Inc., New York; Interscience Publishers Ltd, London.* 1960. *72s. net.*

This collection of twelve articles, to which has been added a bibliography of all his known published work, was originally intended as a sixtieth-birthday gift to Pauli: instead, as a result of his death before it was complete, it became a memorial volume. What emerges from these accounts of twentieth-century physics is that there is scarcely any aspect of it, except perhaps molecular structure, where Pauli's work has not been decisive. It was in the study of fundamental problems that his main interest lay: here, imagination and courage are needed, as well as technical ability. Pauli possessed them all. And the results of this are the Exclusion Principle, the neutrino, the Pauli spin matrices, the general formulation of quantum electrodynamics, and much else. Some of the historical chapters, particularly those by Kronig and Van der Waerden, provide admirable accounts of topics of central importance, and they show, far better than a series of more formal papers, the curious way in which fundamental discovery really proceeds. Most of the book is understandable by any scientist with some training in the physical sciences. The majority of the chapters are in English, but three are in German.

C. A. COULSON

THERMODYNAMICS

Monographs in Statistical Physics and Thermodynamics, *Vol. II*, Thermodynamics, *by P. T. Landsberg. Pp. x+499. Interscience Publishers Inc., New York; Interscience Publishers Ltd, London.* 1961. £5 *9s. net.*

The author of this book is Professor of Applied Mathematics at University College, Cardiff, which helps to explain his approach to thermodynamics as 'among the most abstract branches of Physics'. While this book undoubtedly fulfils the author's aim, which is 'to supplement existing texts, and contribute something to our understanding of thermodynamics', it would not be suitable for a newcomer. Thermodynamics is too useful a science to be omitted, even in elementary courses, but it requires very careful introduction. In this book the treatment is clearly recognized by the author as unconventional. In the first of no fewer than seven appendices, Professor Landsberg attempts to define some of the more unusual features in his book, including the use of the concepts of 'diathermic' and 'adiabatic' partitions. Many interesting, and quite a few important, questions are tucked away in the six other appendices and in the problems presented at the end of each chapter. For advanced reading, the book contains much that gives access in an original way to familiar aspects of thermodynamics, though the degree of originality and unconventionality of treatment are not in all cases justified by the economy of method or profundity of results achieved. A course of reading in this book, for those already well grounded in statistical thermodynamics developed in a more orderly way, is likely to be valuable for theoretical physicists, and would enlarge their experience.

A. R. UBBELOHDE

FLAME PHOTOMETRY

Flame Photometry, *by J. A. Dean. Pp. viii+341. McGraw-Hill Book Co. Inc., New York; McGraw-Hill Publishing Co. Ltd, London.* 1960. *89s. net.*

This book is one of a series on advanced chemistry, and is written primarily for those using flame photometry as an analytical method, for whom it can be strongly recommended.

The data on the various elements, which will be of principal value to the intended reader, are comprehensive both in number and detail. Much of the material is new, and the presentation is systematic and orderly. Of particular interest is Professor Dean's own work on organic-solvent extraction of metals before estimation by flame photometry, and on the increased sensitivity obtained with organic solvents in the flame. The gain in sensitivity is thought to be due partly to the resultant increase in flame size. This points to inadequate filling of the monochromator, but the appropriate remedy is not suggested.

Certain criticisms can be made, mainly in the chapters devoted to instrumentation. Effects attributed to the magnitude of the slit width are somewhat vague. A table (9-1) intended to illustrate the percentage increase in flame background with increased slit width appears to show the converse. Other anomalies tend to mar these sections of the book but are unlikely to perplex the general reader, who should find much of value and interest.

R. L. WARREN

TRANSPORT PHENOMENA

Transport Phenomena, *by R. B. Bird, W. E. Stewart, and E. N. Lightfoot. Pp. xxi+780. John Wiley & Sons Inc., New York and London.* 1960. £5 *10s. net.*

This book is based on notes issued to students in the chemical engineering course at the University of Wisconsin. It deals with the transport of momentum, energy, and mass—that is, matter—in fluids. The analogies between these processes, both in the formal and physical senses, are pressed as far as they will go, although there is of course no analogy in the case of radiative-energy transport, which is also included.

The theme is logically developed and rigorously treated, and in these respects is a long way ahead of most of the texts for chemical engineers in this field. It is quite uncompromising in its approach, and far from easily readable. It would, I think, need the backing of an interpretative set of lectures if used as a text for undergraduates.

The current tendency for the engineering syllabus in U.S. universities to 'go scientific' is here reflected in an avoidance of topics, such as heat transfer to boiling liquids, that cannot be dealt with analytically. I believe this trend may go too far, and that we ought to produce more powerful teaching methods for developing both insight and the qualitative analysis of problems. Meanwhile, this is a first-rate book of its kind.

P. V. DANCKWERTS

AMMONIA SYNTHESIS

La Sintesi dell'Ammoniaca, *by C. A. Vancini. Pp. xxvii+1075. Editore Ulrico Hoepli, Milan.* 1961. *Lire* 12 000.

Total world production of nitrogen compounds, mainly derived from synthetic ammonia, now exceeds twelve million tons of equivalent nitrogen per year, and is still expanding rapidly. Many variations in synthesis-plant design have been introduced since ammonia was first synthesized commercially, and a wide variety of processes are now in use for the production and purification of the vast amounts of hydrogen required. This book sets out the technical and scientific background of these processes in a comprehensive fashion. New processes for hydrogen production, such as steam re-forming of methane and partial oxidation at pressure, are included, as well as the long-established processes, such as those based on coke, coke-oven gas, and electrolysis of water. The recently developed purification processes, such as CO_2 removal with hot potassium carbonate solutions, are included, except, surprisingly, the Italian 'Vetrocoke' process. Low-temperature processes for the production of oxygen, nitrogen, and mixtures of nitrogen and hydrogen are also given. Design of pressure vessels and of ammonia-synthesis plant generally is well covered. Each chapter has a comprehensive bibliography of technical and scientific literature, with, however, few references to patents. The book is well provided with line diagrams, and the theoretical parts are illustrated with simple calculations. Managers and designers of ammonia plants, or those concerned with large-scale hydrogen production, will find it a most useful book. An English translation would be very welcome. T. J. P. PEARCE

ORGANIC ANALYSIS

Organic Electronic Spectral Data. *Vol. I* (1946–52), *edited by M. J. Kamlet, pp. xiv+1208. Vol. II* (1953–55), *edited by H. E. Ungnade, pp. x+919. Vol. I,* £10 15*s. net; vol. II,* £6 12*s. net. Interscience Publishers Inc., New York; Interscience Publishers Ltd, London.* 1960.

The visible and ultra-violet spectra of organic substances have been studied very extensively in recent years. The theoretical interpretation of the spectra, and especially the empirical correlation of the spectra with molecular structure, have been of great value to the organic chemist. The volume of published spectra has grown enormously, as convenient and accurate photoelectric spectrophotometers have become available; a group of over fifty chemists has started to prepare a compilation of all available spectra.

Volumes I and II cover the years 1946–52 and 1953–55, and the editors and their collaborators are to be congratulated on undertaking their enormous task. These two volumes contain a mass of data that will be of great value to all organic chemists.

The compounds are arranged according to molecular formulae, as in 'Chemical Abstracts'. The tables give the compound name, the solvent used, the wavelengths of the absorption maxima, the extinction coefficient of each band, and a reference to the original paper. In Volume I alone the tables cover 1160 pages.

R. E. RICHARDS

QUINONES

Ciba Foundation Symposium on Quinones in Electron Transport, *edited by G. E. W. Wolstenholme and C. M. O'Connor. Pp. xii+*453. *J. & A. Churchill Ltd, London.* 1961. 60*s. net.*

The discovery, by Professor R. A. Morton and his collaborators at Liverpool University, and independently by Professor D. E. Green's group at the University of Wisconsin, of compounds of the ubiquinone (coenzyme Q) series, has provided fresh stimulus to the study of the roles of lipids in electron transport and in the associated phosphorylation processes in mitochondria. We are indebted to the initiators and organizers of this symposium, which took place in 1960, for providing the means whereby representatives of many groups of investigators could report and discuss the initial discoveries, the elucidation of chemical structures, and the many ramifications which are now being pursued.

Apart from authoritative contributions directly related to the title, aspects of the subject which received attention include *in vitro* chemical mechanisms of oxidative phosphorylation; possible roles of quinones (including vitamins K) and of related chroman derivatives in phosphorylation; the significance of plastoquinone (Kofler's quinone) in photosynthetic electron transport and phosphorylation; the biosynthesis of the polyisoprenoid and quinoid portions of the ubiquinones; and incidental critical appraisal of techniques employed in these investigations.

The scope of the symposium demonstrates the attraction of the field for a variety of specialists, and this book will be valuable to an equally wide circle of readers. M. B. THORN

BIOCHEMISTRY

Metabolic Pathways, *Vol. I, edited by D. M. Greenberg* (*second edition of* Chemical Pathways of Metabolism). *Pp. xv+*572. *Academic Press Inc., New York; Academic Press Inc.* (*London*) *Ltd, London.* 1960. £6 1*s.* 6*d. net.*

This book should be of great interest to those—and there must be many—who wish to consult a fully documented reference book that surveys existing knowledge on metabolic pathways in animals, plants, and micro-organisms. This second edition has been virtually rewritten. Of the fourteen contributors to Volume I, only four were among the authors of the first edition. In the seven years since the first publication, knowledge of intermediary metabolism has grown extremely rapidly—many gaps have been filled and many new and unexpected metabolic sequences have been discovered. The subjects covered include free energy and entropy in metabolism (A. B. Pardee and L. L. Ingraham); mitochondrial system of enzymes (D. E. Green and S. Fleischer); glycolysis (B. Axelrod); the tricarboxylic acid cycle (H. A. Krebs and J. M. Lowenstein); other pathways of carbohydrate metabolism (B. Axelrod); metabolism of steroid hormones (L. T. Samuels); and carotenoids and vitamin A (G. Mackinney).

H. A. KREBS

INSECT BIOCHEMISTRY

The Biochemistry of Insects, *by D. Gilmour. Pp. xii+*343. *Academic Press Inc., New York; Academic Books Ltd, London.* 1961. 64*s. net.*

This is the first book to have been written on insect biochemistry and its appearance is timely, because this field has recently become very much more active. The author set out to describe the chemistry of insects as organisms, rather than the chemistry of reactions, and he has achieved this by relating the various biochemical processes to the peculiarities of insect physiology and biology. The result is a very readable book.

Well over half the book is devoted to various aspects of energy production, and in these chapters we see clearly how most of the processes are similar to those of other animals, but with important differences of detail, as in glycolysis and the respiratory chain. The remainder of the book is principally concerned with the intermediate metabolism of carbohydrates, lipids, and nitrogen compounds, including surveys of the components of insect cuticle and of the theories of cuticle tanning.

A number of important papers have been overlooked (for example, the nutrition of mosquitoes is described from early papers, not from the later definitive work), and the references are listed in a manner which could hardly be less convenient for the reader; but the book as a whole provides a most useful critical survey of insect biochemistry. A. N. CLEMENTS

BOTANY

Botanisk Atlas, *Vol. II, by O. Hagerup and V. Peterson (English translation by H. Gilbert Carter). Pp.* 299. *Ejnar Munksgaard, Copenhagen.* 1960. *D.kr.*76.

This is a companion volume to volume 1, by the same authors, which was reviewed in ENDEAVOUR in July 1957. It extends the work to the conifers and vascular cryptogams, and is carried out on the same general plan and with the same merits as its precursor. The omission to state magnifications of the drawings that was noted in the previous volume has been made good in this, to the great increase of their usefulness. About half the Danish species of mosses and liverworts are illustrated, and all those of the lycopods, horsetails, and conifers. As before, there are habit drawings (the trees are represented only by small shoots), usually from fresh material, and enlarged drawings of detail where required. There is rather more accompanying text than in the first volume. It consists mainly of a description of each of the groups illustrated, and is given in Danish and English in parallel columns. Finally there is a section of phylogenetic speculation, with particular reference to the origin of the seed habit, which seems a little at variance with the factual nature of the main body of the work. It is to be regretted that in the illustrations of some groups, such as the Sphagnales, the usual clarity of outline has given place to a fuzziness which leaves a good deal to the imagination. Nevertheless, with the completion of this second volume the authors have increased the debt that botanists owe them. W. O. JAMES

INVERTEBRATE ZOOLOGY

Traité de zoologie, *edited by P.-P. Grassé. Vol. V, Part* 2: Bryozaires, Brachiopodes, Chétognathes, Pogonophores, Mollusques. *Pp.* 1168. *Masson et Cie, Paris.* 1960. *Bound, NFcs* 190; *paper covers, NFcs* 180.

This volume, the second of three concerned with the invertebrates between Platyhelminthes and Nematodes on the one side and Arthropoda on the other, maintains the very high standard of technical excellence set by this magnificent treatise, which now comes near to completion. Both text-figures and plates are models of clarity.

The Bryozoa are very fully covered in a chapter of almost 300 pages by Paul Brien, followed by an account of their fossil representatives by Emil Buge. This is followed by chapters on recent and fossil Brachiopoda, by Paul de Beauchamp and Jean Roger respectively, in which recent work on feeding currents and embryology is well summarized, and the structure, classification, and evolutionary history of the group are thoroughly described. The section on the Chaetognatha, also by Paul de Beauchamp, appears somewhat brief, and certainly much more could have been made of the great ecological importance of these animals in the plankton of all seas.

The chapter of some 100 pages on the relatively recently established group of the Pogonophora is of particular value. It is by A. V. Ivanov of the University of Leningrad, the recognized authority on these animals, and one of the few contributors who is not French. This is the best general account available, and has the basic structure displayed, semi-diagrammatically, in two admirable coloured plates.

The remainder of the volume, somewhat over 500 pages, deals with the Molluscs, excluding Gastropoda, Scaphopoda, and Cephalopoda, which will form the subject matter of the third volume. A. Portmann deals with the general characters of the phylum, and exception might be taken to his failure to note the different symmetries of the body and of the mantle that are the key to an understanding of the wide variety of form within the Mollusca. E. Fischer-Piette and A. Franc deal very adequately with Aplacophora and Polyplacophora, while the Monoplacophora are covered by H. Lemche and K. G. Wingstraad, joint authors of the recent description of *Neopilina*. The value of some of their functional inferences must await examination, unlikely to be long delayed, of living specimens. P. Manigault deals briefly with the shell, and then A. Franc covers the great class of the Bivalvia. Writing as one whose work, with that of co-workers, looms large in this chapter, I can say that modern work has been well summarized and well illustrated. There is certainly no better account available. While the continued use of Pelseneer's classification can hardly be defended, the large systematic section contains excellent descriptions of superfamilies and families. Finally, fossil bivalves are well covered, by Colette Dechaseaux.

All interested in any of these groups should obtain, or make sure they have access to, this fine book.

C. M. YONGE

CERAMIC SCIENCE

Oxide Ceramics, Physical Chemistry and Technology, *by E. Ryshkewitch. Pp. viii* + 472. *Academic Press Inc., New York; Academic Press Inc. (London) Ltd, London.* 1960. $16.

The information this book contains could be presented much more clearly in about one-third of the number of pages used in the present volume. For example, a paragraph of nine lines is found necessary to define Young's modulus, and this paragraph includes the statement 'Perhaps the most significant strength magnitude of a solid body is its Young's modulus of elasticity'. Similar inexactitudes may be found on nearly every page.

The author is not writing in his native language; this book is based on his *Oxydkeramik der Einstoffsysteme*, published in 1948. It is a great pity that someone could not have been found to help improve the style of writing, for there appears to be much useful information to be extracted from this volume. The first sixty-eight pages are a general introduction to oxide ceramics; then follow chapters on alumina (150 pages), spinel (23 pages), magnesia, beryllia, zirconia, and thoria and ceria, and a short section on oxide-metal and oxide-carbide ceramics. The author index is comprehensive and the subject index is adequate. The book appears to be addressed primarily to those concerned with the development of articles made from oxide-ceramics, and as a source of data and a guide to the literature it will be useful. There are many useful hints on the preparation of oxide-ceramic articles; it is in this practical field that the author has made noteworthy contributions.

R. W. DOUGLAS

077826

Short notices of books

(These notices are descriptive rather than critical and are designed to give a general indication of the nature and scope of the books.)

Quantum Mechanics, *by J. L. Powell and B. Crasemann. Pp.* x+495. *Addison-Wesley Publishing Co. Inc., Reading, Mass.* 1961. 74*s. net.*

This is an introduction to quantum mechanics intended for senior students at universities. It aims at emphasizing the physical basis of the subject while not ignoring the mathematics. There are, *inter alia*, chapters on the origins of the quantum theory, the Schrödinger equation, operators and eigenfunctions, matrix mechanics, perturbation theory, and identical particles.

Progress in Low Temperature Physics, *Vol. III, edited by C. J. Gorter. Pp. xii*+495. *North-Holland Publishing Co., Amsterdam.* 1961.

Volume IV, like the first two, is a collection of review articles intended for research workers. Some of the chapters are: vortex lines in liquid helium II, by W. F. Vinen; the nature of the λ-transition in liquid helium, by M. J. Buckingham and W. M. Fairbank; recent developments in superconductivity, by K. W. Taconis; solid-state masers, by N. Bloembergen; and some solid-gas equilibria at low temperatures, by Z. Dokoupil.

A Reference Book of Chemistry, *by J. H. White. Pp.* 302, *University of London Press Ltd, London.* 1960. 21*s. net.*

This book is intended as a supplement to the textbooks used in the higher forms of schools. The sections include: definitions and explanations of important terms, named organic reagents and tests, named inorganic reagents and tests, and named organic reactions. Appendices deal with first aid in the laboratory and with 'useful data'.

Principles of Physical Chemistry, *by W. H. Hamill and R. R. Williams, Jr. Pp. x*+607. *Oliver and Boyd Ltd, Edinburgh.* 1960. 30*s. net.*

The treatment used in this book is to develop thermodynamics first, and then chemical physics; it is intended for use in universities. The whole of fundamental physical chemistry is covered, with chapters on, among other subjects, the first law of thermodynamics, chemical equilibrium in gaseous systems, entropy and the third law of thermodynamics, phase diagrams, nuclear chemistry, and photochemistry and radiation chemistry.

An Introduction to Co-ordination Chemistry, *by D. P. Graddon. Pp. vii*+111. *Pergamon Press Ltd, Oxford.* 1961. 25*s. net.*

The advances of the last two decades are covered by this book, which also includes a brief chapter on the early history of co-ordination chemistry. There are also chapters on polymerization and co-ordinate saturation, the stability of complex salts, and carbonyls and π-complexes. The book is intended primarily for students.

Handbuch der Papierchromatographie, *Vol. II*, Bibliographie und Anwendungen, *edited by I. M. Hais and K. Macek. Pp. xxiv*+726. *Gustav Fischer Verlag, Jena.* 1960. *DM* 44.

This second volume of an extensive treatise on paper chromatography consists of a comprehensive classified bibliography of 10 290 items, referring mainly to the period 1943–56. The book contains, firstly, a general part that covers history, general reviews and books, the theory and techniques of chromatography and preparative chromatography, and, secondly, a special part containing sections on the chromatography of groups of compounds, for example the phenols and related substances, organic acids, steroids. The sections themselves are further subdivided into techniques and applications and are preceded by brief introductory paragraphs of text. The copious indexes fill nearly 200 pages.

Modern Chemical Processes, *Vol. VI, by the editors of 'Industrial and Engineering Chemistry' in conjunction with the technical staffs of the co-operating organizations. Pp.* 126. *Reinhold Publishing Corporation, New York; Chapman and Hall Ltd, London.* 1961. 48*s. net.*

This series aims at describing the processes of American chemical technology in terms of their principles, their plant, their economics, and their future. Some of the techniques described in this volume are: the manufacture of basic silicone products, the use of Grignard reagents, the manufacture of specialized surface-active agents, and the manufacture of acrylates and methacrylates.

A Bibliography of Eastern Asiatic Botany, *by E. H. Walker. Pp. xl*+552. *American Institute of Biological Sciences, Washington, D.C.* 1960. $18.50 *to individuals and industrial libraries,* $16.50 *to AIBS members and all other libraries.*

The original Bibliography, compiled by E. D. Merrill and E. H. Walker, was published in 1938, and covered the literature to the end of 1936. The present volume extends the coverage to the end of 1958 and is arranged in the same way as the original volume. The author explains that Japan and, more markedly, China have not been completely covered, because of the wartime destruction of some libraries in Tokyo and because of the poor circulation of publications in China during the war.

Non-convertible Coatings, *by The Oil and Colour Chemists' Association. Pp.* 326. *Chapman and Hall Ltd, London.* 1961. 30*s. net.*

The primary purpose of this book is to help students to prepare for the Non-convertible Coatings Section of the City and Guilds Examination in Paint Technology, although it is hoped that it will be useful to those in the industry. It is the first of six parts. The contents include chapters on nitrocellulose, plasticizers, lacquers, cellulose acetate, synthetic thermoplastic lacquer resins, and spirit varnishes.

The Major Achievements of Science, *by A. E. E. McKenzie. Vol. I, pp. xvi* +368. *Vol. II, pp. xii*+195. *Cambridge University Press, Cambridge.* 1960. *Vol. I,* 30*s. net (School edition* 20*s. net); Vol. II,* 17*s.* 6*d. net (School edition* 12*s.* 6*d. net).*

The author of this book has aimed at giving students of the humanities a background of science. He gives a survey of scientific discoveries, in chapters such as 'The circulation of the blood', 'Field physics', and 'The germ theory of disease', and of the historical background, in chapters devoted to the seventeenth, eighteenth, nineteenth, and twentieth centuries. The patterns of scientific thought, and the characters of the scientists, are emphasized. The second volume consists of extracts, translated where necessary, from the writings of distinguished scientists.

Notes on contributors

A. F. WAGNER

A.B., M.A., Ph.D.

Was born in Jersey City, New Jersey, in 1922 and was educated at Princeton University, from which he received a Ph.D. degree in 1951. He joined the research staff of Merck & Co. Inc. after leaving the university. His main interests are in the synthesis, and the relationship between structure and activity, of natural products; he participated in the first effective synthesis of mevalonic acid.

K. FOLKERS

B.S., Ph.D.

Was born in Decatur, Illinois, in 1906 and was educated at the University of Illinois, and then at the University of Wisconsin, where he was awarded a Ph.D. in 1931. From 1931 to 1934 he was a Post-Doctorate Research Fellow at Yale University under Professor Treat B. Johnson. He joined Merck & Co. Inc. in 1934; since 1956 he has been Executive Director of Fundamental Research. His interests have been in the chemistry of therapeutic agents, such as alkaloids, vitamin B6, and pantothenic acid; his more recent studies have resulted in the discovery of mevalonic acid, and he has lately been investigating the chemistry of the coenzyme Q group. He has received the Perkin Medal of the Society of Chemical Industry, and is President-elect of the American Chemical Society.

SIR EDWARD BULLARD

M.A., Sc.D., F.R.S.

Was born in 1907 and was educated at Repton School and Clare College, Cambridge. He began research under Lord Rutherford in 1929, working on the scattering of slow electrons in gases. In 1931 he took up geophysics and has done experimental work on gravity, seismology, and the measurement of heat flow on land and at sea. From 1950 to 1955 he was Director of the National Physical Laboratory, and is now Reader in Geophysics in the University of Cambridge. He has written extensively on the origin of the Earth's magnetic field and on magneto-hydrodynamics.

M. H. PIRENNE

M.A., Ph.D., Dr.Sc.

Was born in Verviers, Belgium, in 1912, and was educated at the University of Liège. From 1942 to 1945 he was seconded to Sir F. C. Bartlett's laboratory in Cambridge for research on vision. From 1945 to 1948 he was an I.C.I. Research Fellow at the University of London, from 1948 to 1955 Senior Lecturer in Physiology at the University of Aberdeen. Since 1955 he has been at the University of Oxford, where he is now a University Demonstrator in Physiology. In 1957 he was elected a Vice-President of the *Comité International de Photobiologie*. His main interests in research are the physiology of vision, and in subjects related to this.

V. A. EYLES

D.Sc., F.R.S.E., F.G.S.

Was born in Bristol in 1895 and was educated at Fairfield Secondary School and Bristol University. After service with the Royal Engineers and Royal Air Force (1915–19), he was appointed geologist to the Geological Survey of Great Britain (Scottish Branch) in 1920 and transferred to the English branch as District Geologist in 1945, a post held until his retirement in 1955. He is a founder member of the British Society for the History of Science and member of the History of Science Society (U.S.A.). He has contributed papers on the history of geology to a number of periodicals, and lectured on the subject in the Department of the History and Philosophy of Science, University College, London, from 1951 to 1956, and also at the Universities of California and Illinois.

H. SJÖRS

Fil. dr.

Was born in Sweden in 1915 and was educated at Uppsala University, where he became Lecturer in Plant Ecology in 1948. He later spent three years at the University of Lund, and since 1955 has been a *laborator* in forest botany at the School of Forestry, Stockholm. His main work has been on the ecology of bog and fen vegetation. He has also published works on the phytogeography of Scandinavia, on humus soils, and on northern woodlands conditioned by ancient types of rural management, and is actively interested in nature conservation.

F. F. HEYMANN

B.Sc.(Eng.), Ph.D.(Eng.)

Was born in South Africa in 1924 and was educated at Capetown University from 1940 to 1945. He was on the staff of the University until 1947, when he came to England and joined Metropolitan-Vickers as an industrial research physicist. Since 1950 he has been on the staff of University College, London, where he is a Reader in Physics. He has published extensively on particle accelerators and high-energy nuclear physics; for the last three years he has been engaged in research on positronium.

पुस्तकालय

Some books received

(*Note.* Mention of a book on this page does not preclude subsequent review.)

GENERAL SCIENCE

Art and the Scientist, *by G. Lapage. Pp. xii+115+22 plates. John Wright & Sons Ltd, Bristol.* 1961. 42*s. net.*

Science in Writing, *by T. R. Henn. Pp.* 248. *George G. Harrap & Co. Ltd, London.* 1960. 20*s. net.*

Scientific Thinking and Scientific Writing, *by M. S. Peterson. Pp. vii+*215. *Reinhold Publishing Corporation, New York; Chapman and Hall Ltd, London.* 1961. 56*s. net.*

MATHEMATICS

Boolean Algebra and its Applications, *by J. E. Whitesitt. Pp. x+*182. *Addison-Wesley Publishing Co. Inc., Reading, Mass., and London.* 1961. 51*s. net.*

Differential Equations, *by F. G. Tricomi (translated from the Italian by Elizabeth A. McHarg). Pp. x+*273. *Blackie & Son Ltd, London.* 1961. 50*s. net.*

Measure, Lebesgue Integrals, and Hilbert Space, *by A. N. Kolmogorov and S. V. Fomin (translated by N. A. Brunswick and A. Jeffrey). Pp. xii+*147. *Academic Press Inc., New York; Academic Press Inc. (London) Ltd, London.* 1961. $4.

ASTRONOMY

Atoms to Galaxies, an Introduction to Modern Astronomy, *by J. Stokley. Pp. vii+*361. *The Ronald Press Co., New York.* 1961. $6.

Frontiers of Astronomy, *by F. Hoyle. Pp. xvi+*360. *William Heinemann Ltd, London,* 1955; *Mercury Books, London,* 1961. 12*s.* 6*d. net.*

Introduction à l'astrophysique: les étoiles, *by J. Dufay. Pp.* 218. *Collection Armand Colin, Paris.* 1961. *NFcs* 4.50.

PHYSICS

The Birth of a New Physics, *by I. B. Cohen. Pp.* 200. *William Heinemann Ltd, London.* 1961. 5*s. net.*

From Dualism to Unity in Quantum Physics, *by A. Landé. Pp. xvi+*114. *Cambridge University Press, London.* 1960. 18*s.* 6*d. net.*

The Physics of Television, *by D. G. Fink and D. M. Lutyens. Pp.* 160. *William Heinemann Ltd, London.* 1961. 5*s. net.*

Principes de radiocristallographie, *by J. Barraud. Pp. xii+*236. *Masson et Cie, Paris.* 1960. *NFcs* 42.

Problems in Quantum Mechanics, *by I. I. Goĭldman and V. D. Krivchenkov, edited by B. T. Geĭlikman (translated from the Russian by E. Marquit and E. Lepa) (revised edition). Pp. viii+*275. *Pergamon Press Ltd, Oxford.* 1961. 50*s. net.*

CHEMISTRY

Chimie générale et abrégé de chimie minérale, *by P. Colmant. Pp. xv+*565. *Masson et Cie, Paris.* 1960. *NFcs* 45.

BIOLOGY

Biological Problems arising from the Control of Pests and Diseases, *edited by R. K. S. Wood. Pp. viii+*120. *The Institute of Biology, London.* 1960. 25*s. net.*

Electron Microscopy, *by E. H. Mercer and M. S. C. Birbeck. Pp. vii+*76. *Blackwell Scientific Publications Ltd, Oxford.* 1961. 9*s.* 6*d. net.*

Evolution after Darwin, *edited by Sol Tax, Vol. I,* The Evolution of Life: Its Origin, History and Future. *Pp. viii+*629. *Vol. II,* The Evolution of Man: Mind, Culture and Society. *Pp. viii+*473. *Chicago University Press; Cambridge University Press, London.* 1960. 80*s. net each.*

Journal of Theoretical Biology, *Vol. I, No.* 1, January 1961, *editor in chief: J. F. Danielli. Pp.* 106. *Academic Press Inc. (London) Ltd, London.* 1961. £6 1*s.* 6*d. net the volume; private subscriptions* 86*s. net.*

Problems of the Physiology of the Processes of Fatigue and Recovery *(translated from the Russian by M. Roublev). Pp.* 264. *Program for Scientific Translations, Jerusalem.* 1960.

The Processes of Ongoing Human Evolution, *edited by G. W. Lasker. Pp.* 108. *Wayne State University Press, Detroit.* 1960. $3.75.

ZOOLOGY

Les écrevisses françaises, *by M. André. Pp. vi+*293. *Editions Paul Lechevalier, Paris.* 1960. *NFcs* 18.00.

MEDICINE

Journal of Atherosclerosis Research, *Vol. I, No.* 1, January/February 1961. *Pp.* 92. *Elsevier Publishing Co., Amsterdam.* 1961. £6 5*s. net per year.*

Problems of Infection, Immunity and Allergy in Acute Radiation Diseases, *by N. N. Klemparskaya, O. G. Alekseyeva, R. V. Petrov, and V. F. Sosova (translated by Lydia Venters). Pp. viii+*165. *Pergamon Press Ltd, Oxford.* 1961. 50*s. net.*

Surgery is destined to the Practice of Medicine. Ninetieth Hunterian Oration, 1959, *by R. Watson-Jones. Pp.* 81. *E. & S. Livingstone Ltd, Edinburgh.* 1961. 21*s. net.*

ENGINEERING

Proceedings of the Symposium on Active Networks and Feedback Systems, New York, April 1960, *edited by J. Fox. Pp. xxv+*658. *Polytechnic Press of the Polytechnic Institute of Brooklyn; Interscience Publishers Inc., New York; Interscience Publishers Ltd, London.* 1961. $8.

TECHNOLOGY

Books on the Chemical and Allied Industries: A Subject Catalogue of Books in the Science Library, *compiled by L. R. Day. Pp. ix+*118. *Her Majesty's Stationery Office, London.* 1961. 12*s.* 6*d. net.*

Chemical Processing of Fuel. Proceedings of the second All-Union Conference on Artificial Liquid Fuel and Gases (selected papers from) *(translated from the Russian). Pp.* 118. *Program for Scientific Translations, Jerusalem.* 1961. 35*s. net.*

Fuel Research, 1917–1958. *Pp. viii+*120. *Department of Scientific and Industrial Research; Her Majesty's Stationery Office, London.* 1961. 15*s. net.*

Water Pollution Research, 1960. *Pp. vi+*122. *Department of Scientific and Industrial Research; Her Majesty's Stationery Office, London.* 1961. 7*s. net.*

BIOGRAPHY

American Men of Science: a Biographical Directory, *edited by J. Cattell (tenth edition).* The Physical and Biological Sciences, F–K. *Pp.* 1127–2284. *The Jacques Cattell Press Inc., Temple, Arizona.* 1960. £9 12s. 6d.

HISTORY OF SCIENCE

Die Naturwissenschaftler und ihre Philosophie. Geistesgeschichte der Chemie, *by W. Böhm. Pp. xvi+*316. *Herder, Vienna.* 1961. 120 *Schillings.*

Made and printed in Great Britain at THE KYNOCH PRESS, *Witton, Birmingham,* 6, *for the publishers*
IMPERIAL CHEMICAL INDUSTRIES LIMITED, *London, S.W.*1

Compared
1999-2000